Systematic Reviews in the Field of Complementary and Alternative Medicine:

Importance, Methods and Examples Concerning Acupuncture

Eric Manheimer

Systematic reviews in the field of complementary and alternative medicine: importance, methods and examples concerning acupuncture

The studies presented in this thesis were partially funded by grant number R24 AT001293 from the National Center for Complementary and Alternative Medicine (NCCAM) of the US National Institutes of Health (www.nccam.nih.gov). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NCCAM, or the US National Institutes of Health.

ISBN: 978-0-615-87496-8

Art work credits: Amanda Johnson
"Wild Card" 2013
AmandaJohnson@gmail.com

Printing: VU Copy Department

© Copyright Eric W. Manheimer, 2013
No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without the permission in writing of the author, or, when appropriate, of the publishers of the publications.

VRIJE UNIVERSITEIT

Systematic reviews in the field of complementary and alternative medicine: importance, methods and examples concerning acupuncture

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 7 november 2013 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Eric William Manheimer

geboren te Boston, MA

promotoren:	prof.dr. L.M. Bouter
	prof.dr. D.A.W.M. van der Windt
copromotor:	prof. B. Berman, MD

Table of Contents*

Chapter 1 7
General Introduction

Chapter 2 19
Development and classification of an operational definition of complementary and alternative medicine for the Cochrane Collaboration

Chapter 3 37
Bibliometric analysis of the Cochrane Complementary Medicine Field specialized register of controlled trials

Chapter 4 55
Evidence from the Cochrane Collaboration for traditional Chinese medicine therapies

Chapter 5 75
Systematic reviews and CAM

Chapter 6 89
Published reports of acupuncture trials showed important limitations

Chapter 7 101
Meta-analysis: acupuncture for low back pain

Chapter 8 119
Acupuncture for peripheral joint osteoarthritis: Cochrane review

Chapter 9 151
The effects of acupuncture on rates of clinical pregnancy among women undergoing in vitro fertilization: systematic review and meta-analysis

Chapter 10 179
General Discussion

Appendix. Definitions of abbreviations and acronyms 209
Summary 211
Samenvatting 217
Acknowledgments 223
About the author 229
List of publications 231

* An extended, online version of the Table of Contents can be accessed by using the 'PhD thesis' link on my personal webpage, available at https://sites.google.com/site/manheimereric/home. This extended, online Table of Contents includes links to the full-text of the following publications: each thesis chapter together with any supplementary appendices (for Chapters 3, 7, and 9); preliminary, preparatory work published as either abstracts or journal articles (for Chapters 3, 4, and 9); earlier versions of the systematic reviews on acupuncture for osteoarthritis (Chapter 8) and IVF (Chapters 9); the complete, extended publication of the osteoarthritis Cochrane review (Chapter 8), which has been substantially abridged for inclusion in this print thesis; and a relevant Letter to the Editor and our response (for Chapter 7). The full-text of each of these documents will open when the relevant item is clicked on the online Table of Contents page. In this print version of the thesis, the relevant, supplementary online material for each chapter is listed at the bottom of the chapter's abstract page.

Letter of artistic intent for “Wild Card” cover painting:

“My objective as the artist was to create a modern atmospheric visual 2D plane that captures human cells in change. I used color, composition, and shape to support my objective. The vertical and horizontal lines represent acupuncture tools, the star like shapes represent methods and examples concerning acupuncture. My goal was to support the content of Mr. Manheimer’s thesis with a cover that is colorful, attractive and symbolic of its content.”

Thank you,
Amanda Johnson

Chapter 1. General Introduction

Complementary and Alternative Medicine

Complementary and alternative medicine (CAM) has been defined as a "group of diverse medical and healthcare systems, practices and products that are not presently considered to be part of conventional medicine."[1] Because CAM is defined in relation to conventional medicine, it is important to point out that, in the United States, "conventional medicine (also called Western or allopathic medicine) [has been defined as] medicine as practiced by holders of M.D. (medical doctor) and D.O. (doctor of osteopathic medicine) degrees and by allied health professionals, such as physiotherapists, psychologists, and registered nurses."[1] The boundaries between CAM and conventional medicine are not always clear-cut, may change over time, and may be dependent on geographical location. For example, while osteopathic medicine was founded in the late nineteenth century as a form of holistic medicine that relied largely on the manipulation of bones and joints, by the middle of the twentieth century, the osteopathic profession had moved closer to mainstream medicine, and US osteopathic physicians currently use all conventional methods of diagnosis and treatment, and have the same full practice rights as M.D.s, in all fifty US states.[2] However, while osteopathy is now largely viewed as part of conventional medicine in the United States, it is still perceived as CAM in most European countries, including the United Kingdom.[3]

Some interventions, for instance acupuncture, would be accepted as CAM by virtually everyone. Indeed, acupuncture is always included among the CAM interventions assessed in population surveys of CAM use,[4-8] and one survey[9] identified acupuncture as the best known CAM intervention.[10] However, while acupuncture is considered a CAM intervention, at least in Western countries, it is often provided by practitioners of conventional medicine (e.g., physiotherapists or M.D.s). For example, acupuncture is delivered by 6,500 physiotherapists in the United Kingdom,[11] and by about 40,000 conventionally trained physicians in Germany.[12] Other conventionally trained practitioners often collaborate with or refer patients to acupuncturists. Thus, evidence on the effectiveness of acupuncture would be highly relevant not only to CAM practitioners, but also to conventionally trained practitioners. Patients, meanwhile, are less concerned with the boundaries between what is considered CAM versus conventional medicine, and often seek an integrated approach that incorporates both conventional and complementary elements.[13]

Systematic Review Methodology and Examples Concerning Acupuncture

This thesis focuses on the methodology of systematic reviews because of their importance and relevance to healthcare researchers, providers, consumers, and policy-makers. This thesis has a specific focus on systematic reviews of acupuncture, as well as on research that is relevant for the methods of these acupuncture systematic reviews. It also includes an analysis of the contents of CAM trial and review evidence bases, as developed by the CAM Field of the Cochrane Collaboration.

The example systematic reviews in this thesis concern acupuncture for several reasons. First, acupuncture is used in practice on a very large scale. Second, acupuncture has been evaluated in a large number of randomized controlled trials (RCTs). Indeed, there are an estimated 5,321 trials of acupuncture in the Cochrane Central Register of Controlled Trials (CENTRAL), as of Issue 12, 2012.[14] While probably only a small proportion of these acupuncture trials are of sufficient quality to support evidence-based decision-making, for some health condition topics, the existing acupuncture trials provide a solid evidence base for systematic reviews. Third, acupuncture is a relatively safe intervention, with a very low incidence rate of major adverse effects, according to four large prospective surveys.[12,15-17] Fourth, basic research has suggested plausible putative mechanisms of effect of acupuncture, particularly for pain-related conditions,[18] and this basic research can serve to understand the results of systematic reviews of acupuncture.

High Prevalence of Use and Public Expenditures on CAM Interventions Justify Studies of the Evidence of their Efficacy

Population surveys have shown that the use of CAM by the Public is a common phenomenon in the United States. For example, a 1997 US survey found that 42% of respondents had used at least one CAM intervention in the previous year.[5] Five years later, in 2002, the US Centers for Disease Prevention and Control, National Health Interview Survey (NHIS) found that the proportion had increased to 62%.[19] The most recent US NHIS, conducted in 2007, found that 38% of respondents had used at least one CAM intervention in the previous year.[4] Differences across surveys in these US prevalence estimates of overall use of "CAM" can be largely explained by different definitions of "CAM" in the various surveys, as reflected by differences in the specific CAM interventions that survey respondents were asked about.[4]

Prevalence data of CAM use in other countries is sparse and inconsistent, with the exception of the United Kingdom. Four UK population-based large surveys using random sampling methods have estimated one-year prevalence rates of CAM use to be 10%,[20] 20%,[21] 26%,[22] and 28%.[8] A 2013 systematic review of *all* UK surveys[23] has estimated that the average one-year prevalence rate of CAM use to be 41%; however, this systematic review found that the prevalence rates were lower in the 'higher quality' surveys, where the criteria for 'higher quality' surveys included sample size of >1000, response rate >70%, and the use of random-sampling technique. This 2013 systematic review of UK surveys[23] also illustrates how lower quality surveys may result in inflated estimates of CAM use. Namely, this review found that there was an inverse association between survey response rate and prevalence of CAM use, and that surveys with response rates between 21% and 50% reported prevalence rates that were three times greater than surveys with response rates >70%. The review authors postulated that the reason surveys with low response rates might generate falsely positive prevalence rates is because the survey recipients who do not respond to the survey are probably less interested in CAM, and would therefore be less likely to use CAM, than those who do respond.[23]

A systematic review of surveys of the prevalence of CAM use among general populations in countries worldwide found that a substantial proportion of the populations in the countries surveyed used CAM. However, comparisons, both across countries and within countries, were again difficult because of differences in definitions of CAM, differences in the reference time period for the use of these CAM interventions, differences in survey designs (e.g., postal versus interview), and other differences in the methodology and quality of the surveys.[10,24]

In terms of Public expenditures on CAM, the US NHIS found that US adults spent $33.9 billion total out of pocket on the purchase of CAM products, classes, and materials for self-care, and on visits to CAM practitioners during 2007.[7] This equates to 1.5% of total US health care expenditures and 11.2% of total out-of-pocket costs. Approximately two-thirds of these CAM out-of-pocket costs were spent for purchasing CAM products (e.g., dietary supplements), classes, and materials for self-care, and the remaining one-third for visits to practitioners (e.g., acupuncturists). In terms of out-of-pocket dollars spent on CAM dietary supplements relative to conventional medicine pharmaceutical drugs, the US Public spent $14.8 billion on dietary supplements compared to $47.6 billion on pharmaceuticals in 2007. In terms of out-of-pocket dollars spent on visits to CAM relative to conventional practitioners, the US Public spent $12.4 billion on CAM practitioner visits compared to $49.6 billion spent on conventional physician services.[7]

As of 2007, acupuncturists were the third most commonly visited CAM practitioners in the United States, with massage therapists first, and chiropractors second.[7] However, while the number of visits US adults made to CAM practitioners substantially decreased overall between 1997 and 2007, the number of visits to acupuncturists increased three-fold during this same time period.[7] This increase has been postulated[4] to be partially due to increased awareness of acupuncture from

articles in the lay press extolling its benefits. Despite this increase, current rates of acupuncture use in the US general population are still relatively low, with only about 1.4% of adults having used acupuncture in the past year, according to the most recent US population-wide survey.[4] In Germany and the United Kingdom, the rates are higher, with 9%[25] and 3%[22] of the populations respectively having used acupuncture in the past year, according to national population-based surveys. In the specific subgroup of Chinese immigrants to the United States, 14% of surveyed participants had used acupuncture in the past year.[26] Rates of acupuncture use among people in Asian countries, such as Japan[27] and Taiwan,[28] are substantially higher than rates of use in most Western countries. However, the rates of use of acupuncture in the past year in one European country (i.e., Germany[25]) were higher than the rates of use in Japan (7%[27]) and Taiwan (6%[28]), despite the fact that acupuncture originated in China and is widespread throughout Asia.

Thus, surveys suggest a substantial prevalence of CAM use, and large out-of-pocket expenditures on CAM products and CAM practitioner visits. While there is a fair amount of evidence available on acupuncture (and CAM interventions more generally), for many health conditions, often no firm conclusions for or against acupuncture efficacy can be drawn. This is both because of a shortage of high quality trials in some areas, and also because of the differing conclusions that can be drawn, even from the high quality trials, depending on whether decision-makers focus on the comparison of acupuncture versus sham acupuncture or the comparison of acupuncture versus standard care therapies.[29,30] Indeed, the 2007 NHIS survey authors[4] noted the disparity between the increasing use of acupuncture and the lack of conclusive evidence of efficacy according to systematic reviews, except for treating a few specific conditions. The need for evidence for or against efficacy is great, not only among people with serious, chronic health conditions seeking to relieve symptoms and to slow disease progression, but also among healthcare providers and guideline committees, who require high quality and consistent clinical trial evidence before recommending a medical intervention, especially in this era of evidence-based medicine.

Systematic Reviews: The Most Reliable Method for Synthesizing Evidence on Efficacy of Healthcare Interventions

The systematic review is widely accepted as the most valid and reliable method for evaluating healthcare interventions, including CAM interventions. Systematic reviews use explicit, transparent, and well-documented methods to find, evaluate, and synthesize the best available studies related to a specific research question. Systematic reviews of healthcare interventions have typically focused on RCTs, because RCTs are widely regarded as the study design providing the most valid estimates of a healthcare intervention's effects. Systematic reviewers aim to evaluate and appraise relevant RCTs using objective, transparent, and reproducible methods in order to provide an unbiased assessment of the evidence on the effects of a given intervention. Systematic reviews often include a meta-analysis, the quantitative combining (pooling) of results from similar but separate RCTs with the aim to obtain a valid and reliable pooled effect estimate.

Systematic reviews are complex undertakings. The complexity often arises from the large heterogeneity of objectives, methods and results of RCTs included in systematic reviews, which makes it difficult to summarize the available RCTs evaluating an intervention. Indeed, systematic reviews can rarely draw general, global conclusions that an intervention 'works', but rather estimate the efficacy of a specific intervention relative to a specific type of control (e.g., a placebo, another active treatment, or a waiting list control). This certainly holds for systematic reviews of CAM interventions, as many CAM-related RCTs compare a CAM intervention to one or more control interventions.

Systematic reviews are considered rigorous scientific studies, comparable in scope and complexity to other types of research studies such as RCTs.[31] Indeed, as evidence of its importance,

a citation analysis study found that systematic reviews that include meta-analyses now have the greatest citation impact of all study designs (exceeding even RCTs), and this citation impact is continuing to increase.[32] Systematic reviews also have been shown to receive double the number of citations compared with non-systematic ('narrative') reviews.[33] This citation impact of systematic reviews is also commensurate with its position at the top of the hierarchy of research evidence.[34]

Systematic Reviews: Relevance to Researchers

Systematic reviews can help researchers pinpoint where knowledge gaps exist and thereby help in the design and conduct of new RCTs. The systematic review serves to ensure that a proposed RCT is relevant, necessary, and guided by earlier RCTs. Amidst the vast number of research questions that remain to be addressed in CAM, and the limited financial support available to conduct RCTs of CAM interventions, it is important that CAM investigators plan RCTs in the context of what is already known on a topic, as summarized in a systematic review.

Indeed, systematic reviews have an iterative relationship with RCTs. As an example, the findings of a 2001 systematic review of acupuncture for osteoarthritis[35] guided the design of the large, sham-controlled RCT of this topic, conducted at the University of Maryland School of Medicine, and published in 2004.[36] This RCT, in turn, was included in our 2010 Cochrane review of acupuncture for osteoarthritis (Chapter 8).[37] Our 2010 Cochrane review includes an entire section on Implications for Research, which includes suggestions for comparisons to address in future RCTs, methodological issues to consider (primarily related to blinding), and outcome measures and treatment protocols to use. Thus, this 2010 Cochrane review may consequently help RCT researchers optimally design future acupuncture for osteoarthritis RCTs, and these future RCTs can then be incorporated into the next update of this Cochrane review. Indeed, to insure that future RCTs are optimally designed, based on what has been learned from a systematic review of previous RCTs,[38,39] some funding agencies such as the UK Medical Research Council,[40] and medical journals including the *Lancet*,[41] now require evidence from a recent systematic review before funding or publishing a new RCT.

Systematic Reviews: Relevance to Healthcare Consumers, Providers, and Policy-makers

Up-to-date systematic reviews are also important to healthcare consumers (i.e., the Public). With the advent of the Internet, the Public has increased access to a substantial part of the knowledge base of medicine. While many healthcare journals do not yet provide the Public with free, full open access to their contents, research articles published in these journals are increasingly becoming more available to the Public. For example, recent US regulations require that all peer-reviewed research articles funded by NIH are made freely available online after one year.[42] The complete Cochrane reviews are not in the open access domain, except for among individuals living in countries with a national license, including the United Kingdom, Australia, Denmark, India, and several other countries/regions.[43] However, brief plain language summaries of Cochrane reviews are freely available, and these short summaries may be the most important source of information for the Public. The proportion of Cochrane reviews that are specific to CAM may not be as high as expected, given the high prevalence of CAM use among the Public. That is, of the 5,352 complete Cochrane reviews included in Issue 12, 2012 of *The Cochrane Library*, only about 11% (i.e., 596) relate to CAM.

Providers must also keep current with systematic review findings or they may risk adverse consequences for patient care, including the continued recommendation and use of interventions proven ineffective or even harmful by RCTs, as well as a delay in the uptake of interventions with the potential for benefit in specific populations.[44,45] Similarly, professional organizations need access to systematic reviews (and their summaries) as a way of summarizing evidence for the

development of guidelines, and policy-makers or commissioners use systematic reviews to aid reimbursement decisions.[46]

Systematic Reviews: The Building Blocks for Summary of Findings Tables, Plain Language Summaries, and Evidence-based Clinical Practice Guidelines

Complete systematic reviews are typically long and complex scientific documents. Because most stakeholders do not have the time or skills to read and understand the complete reviews, or do not have access to them, the information from systematic reviews is also incorporated into Summary of Findings tables, Plain Language Summaries, and evidence-based clinical practice guidelines. Summary of Findings tables focus on the major comparison from a Cochrane review (e.g., acupuncture versus sham acupuncture) and show in a tabular format the findings for each outcome for this comparison, as well as an evaluation of the overall strength of the evidence for each outcome. Plain Language Summaries, which can be useful to both the Public and practitioners, are prepared based on the data from the Summary of Findings tables. An example of a Summary of Findings table and a Plain Language Summary is included in the 2010 Cochrane review of acupuncture for osteoarthritis (Chapter 8). In his role as the administrator of the Cochrane CAM Field, the author of this thesis has coordinated collaboration between the Cochrane CAM Field and the Norwegian branch of the Nordic Cochrane Centre to prepare and disseminate Summaries of Findings tables and Plain Language Summaries of CAM-related Cochrane reviews. So far, this collaboration has resulted in the preparation of Summary of Findings tables and Plain Language Summaries for 26 Cochrane reviews, nine of which have been published (or are in press) in Cochrane Summary of Findings columns in the CAM journals *Alternative Therapies in Health and Medicine* and *Global Advances in Health and Medicine*. All of these CAM-related Summary of Findings tables and Plain Language Summaries, as well as their associated journal columns, are freely available on the Cochrane CAM Field website.[47,48]

Systematically derived evidence-based clinical practice guidelines, which are often the basis for healthcare decision-making, use evidence from multiple systematic reviews (among other sources), for the various interventions for a given healthcare condition. Such guidelines often include a short (e.g., 2-3 page) Summary of Recommendations, which provides the busy clinician with quick access to the bottom-line evidence. Evidence-based clinical practice guidelines exist for both osteoarthritis and low back pain, health conditions addressed in two of the systematic reviews of acupuncture included in this thesis (i.e., Chapters 7 and 8). Clinical practice guidelines for osteoarthritis, which include the OARSI,[49] EULAR,[50] the American Academy of Orthopaedic Surgeons,[51] and the UK National Institute for Health and Clinical Excellence (NICE)[52] guidelines do not endorse the use of acupuncture for osteoarthritis, primarily because there is not sufficient, conclusive evidence of clinical effectiveness or cost-effectiveness to allow for a firm recommendation.[51,52] However, these guidelines are periodically updated, and their recommendations can change with changes in the evidence landscape. In contrast, guidelines for low back pain, which include the UK NICE,[53] the American Pain Society/American College of Physicians joint clinical practice guideline,[54,55] and the North American Spine Society guideline[56] generally recommend that clinicians consider acupuncture as one possible treatment option for patients with low back pain. To illustrate how clinical practice guidelines can form the basis for health policy decisions, as a result of the UK NICE recommending acupuncture as a treatment option for patients with low back pain, the UK National Health Service now provides (a maximum of) 10 sessions of acupuncture over a period of 12 weeks for people with low back pain that has persisted for more than six weeks.

The Cochrane Collaboration, *The Cochrane Library*, and the Regularly Updated Cochrane Reviews

The Cochrane Collaboration is an internationally renowned non-profit initiative dedicated to help health care providers, policy-makers, patients, their advocates and carers, make well-informed decisions about health care, based on the best available research evidence, by preparing, maintaining, and promoting the accessibility of systematic reviews. The Cochrane Collaboration currently involves more than 28,000 contributors from 100 countries.[57] Approximately 20% of all systematic reviews appearing each year are published by the Cochrane Collaboration.[58]

The main product of the Cochrane Collaboration is the *Cochrane Database of Systematic Reviews* published as part of *The Cochrane Library*. Cochrane reviews are regularly updated to take into account any eligible new trials. Because the RCT evidence base can change so rapidly, the regular updating of Cochrane reviews is critical for ensuring that *The Cochrane Library* remains an up-to-date and accurate source of the effects of healthcare interventions. Regular updating is made possible through the electronic publication of *The Cochrane Library.* Finally, it is important to note that although the Cochrane Collaboration is widely regarded as setting the standard for excellence in systematic reviews, and *The Cochrane Library* remains the premier independent, high quality evidence base for healthcare decision-making, high quality systematic reviews are also regularly published outside of *The Cochrane Library*.

Outline of the Thesis

The chapters that comprise this thesis are divided into two Parts because the component chapters address two discrete but related key objectives, both pertaining to systematic reviews of CAM interventions. That is, the key objective of Part 1 (Chapters 2-4[59-61]) has been to develop and analyze CAM systematic review-related databases. The key objective of Part 2 (Chapters 5-9[37,62-65]) has been to prepare systematic reviews to evaluate the effects of acupuncture for three specific health conditions, as well as to conduct research with a view to improve the methods of future acupuncture RCTs and systematic reviews.

Part 1. Development and Analysis of CAM Systematic Review-related Databases (Chapters 2-4)

Chapter 2: Development of an operational definition of CAM, for use in creating a database of CAM-related Cochrane reviews.

Chapter 2 of this thesis describes the development of an operational definition of CAM, which was used for creating a database that includes all Cochrane reviews of CAM interventions (i.e., a Cochrane CAM Field "topics list"). The initial motivation for developing this CAM Field topics list was a request from Sir Iain Chalmers, one of the founders of the Cochrane Collaboration, for a regularly updated list of CAM-related Cochrane reviews that he could show in his presentations about the Cochrane Collaboration. At the time of this request, prior to the development of this topics list, we found that each time we tried to generate a list of CAM-related Cochrane reviews by classifying Cochrane reviews as CAM or not CAM we ended up with overlapping but slightly different sets of reviews. The objective of developing the CAM Field topics list was to allow us to reproducibly and consistently classify Cochrane reviews as CAM/not CAM, and thereby reliably track the total number of CAM-related Cochrane reviews in *The Cochrane Library,* as well as the number of Cochrane reviews on specific CAM interventions.

Chapter 3: Bibliometric analysis of a database of CAM-related trials.

A comprehensive database of CAM-related trials is a vitally important resource for CAM systematic reviewers because it could serve as a 'one-stop shopping' source for identifying eligible trials in the

future. In the absence of a comprehensive database of CAM-related trials, reviewers attempting to locate all relevant trials would have to undertake laborious, time-consuming, and resource-intensive searches of electronic databases, page-by-page 'handsearches' of non-indexed medical journals, and scans of reference lists of related reports. Though onerous, such searches are necessary to identify all clinical trials eligible for a review. A failure to identify relevant trials calls into question the findings of a review, especially if the unidentified trials could later be shown to have different results than those of the identified trials. The ultimate aim of developing the CAM Field specialized database of trials is to provide the most comprehensive source of relevant CAM trials available, thus reducing the Cochrane reviewers' burden of searching multiple sources. Chapter 3 describes the methods of developing this database of CAM-related trials, the results of an analysis of its current composition, and the implications of the use of this CAM trials database by systematic reviewers.

Chapter 4: Reviewing the evidence from a database of traditional Chinese medicine-related Cochrane reviews.

Traditional Chinese medicine (TCM) treatments – both acupuncture and Chinese herbal medicine – are widely used therapies and comprise a large proportion of the reviews on the Cochrane CAM Field topics list. The objective of Chapter 4 was to conduct a review overview of the evidence base for TCM according to Cochrane reviews, and to identify directions for future TCM trials and reviews. This chapter also involved evaluating whether or not each of the 70 TCM-related Cochrane reviews suggested a beneficial effect. Because there are no widely accepted methods for preparing an overview of the bottom-line findings of a large number of systematic reviews, developing such a method, and evaluating its strengths and limitations, was a primary challenge of this chapter.

Part 2. Methodological Research on CAM Systematic Reviews (Chapters 5 and 6) and Preparation of Systematic Reviews of Acupuncture for Three Different Health Conditions (Chapters 7-9)

Chapter 5: Review of systematic review methodology in CAM.

Systematic reviews, like all other research study designs, do have methodological challenges and limitations. The objectives of Chapter 5, a methodological review book chapter, were to provide an overview of systematic review methods in relation to CAM, to summarize current methodological research related to the preparation of CAM-related systematic reviews, and to illustrate through case examples (many selected from two of the other chapters included in this thesis[63,65]) various approaches used to address methodological challenges in CAM reviews.

Chapter 6: Comparing trial publications and surveys of the trials' authors to investigate completeness and accuracy of reporting on the methodology of randomization and blinding.

One of the major methodological limitations of systematic reviews is the incomplete reporting of the methodological details about how the included RCTs were conducted. In systematic reviews, the quality of the methods used in the included RCTs is generally assessed based on the information from the RCT publication. A major limitation of assessing quality of methods (most importantly, risk of bias) based solely on the RCT publication is that this relies on a critical assumption that what is written in the publication reflects actual study procedure and that "if it is not reported, it probably was not done." The objective of Chapter 6 was to survey acupuncture RCT authors to determine whether the description of the randomization and blinding methods used in their acupuncture RCT publications was an accurate and complete reflection of the study procedures used.

Chapters 7-9: Systematic reviews of acupuncture for three specific health conditions.
Chapters 7-9 are systematic reviews of acupuncture for the following three health conditions: low back pain, osteoarthritis, and as an adjuvant to in vitro fertilization. Given that acupuncture is used to treat many different conditions, why were these three health conditions prioritized and selected? Indeed, the issue of prioritizing and selecting topics for systematic reviews has been a topic of recent discussion within the Cochrane Collaboration. Some *general* methods for prioritization that Cochrane Review Groups have proposed include a consideration of the prevalence of use of the intervention, the importance of the topic (e.g., cost, burden of disease, public interest, WHO Millennium Development Goals), and the topic controversy.[66,67] For CAM/acupuncture *specifically*, there are few high quality RCTs for many conditions. In the absence of high quality RCTs, systematic reviews cannot draw any conclusions about the intervention's effects. Thus, the availability of a sufficient number of high quality RCTs is another important consideration for prioritizing CAM/acupuncture systematic reviews. The three topics that were selected for the included systematic reviews each met most of the prioritization criteria described above, including relatively large numbers of available RCTs, high burdens of disease, moderately high prevalence of use, and topic controversy (particularly for the IVF review). In addition, for each of the three conditions studied, there are few safe and effective conventional treatments, thereby providing a further "alibi" for evaluating the effects of acupuncture.

Among these three included systematic reviews, two (Chapters 8 and 9) were review *updates*. In discussing priorities, it is also important to stress that *updates* of existing Cochrane reviews, particularly reviews of topics including many RCTs, might be given equal, or indeed, even higher priority to *new* Cochrane reviews. This is because, compared with updates of already existing systematic reviews, new topics not yet covered by systematic reviews may be more likely to include only small, heterogeneous, methodologically unsound trials, and/or to review interventions with a lower prevalence of use for health conditions that have a smaller burden of disease. One important reason for deciding *when* to update a review is whether new RCTs have been published. For these two reviews (Chapters 8 and 9), there was an a priori awareness of many new RCTs, which justified the preparation of updates. More scientific and systematic approaches are also being developed for prioritizing review updates, including an innovative "signal detection method".[68] There are plans to use this method in the future to detect CAM-related Cochrane reviews needing updates.

A primary objective of these three systematic reviews was to estimate the effects of acupuncture for each of the given health conditions, in order to guide evidence-based decision-making. Another objective was to provide recommendations for the design of future RCTs based on a thorough assessment of the strengths and limitations of the existing RCTs. As discussed in the General Discussion Chapter, there were many challenges, both unique to individual reviews and in common across all three reviews that impeded our ability to draw definitive conclusions about acupuncture for the given conditions. However, future updates of these reviews, which will likely include new RCTs that are better designed for accurately estimating treatment effects, may allow more definitive conclusions to be drawn.

Chapter 10: General Discussion
Chapter 10, the last chapter of this thesis, provides a general discussion of the overall findings of the above studies. This General Discussion Chapter also outlines some of the challenges and methodological issues faced in conducting these studies, and the implications of these studies for practice and for future research.

REFERENCES

1. National Center for Complementary and Alternative Medicine. What is complementary and alternative medicine? Bethesda, MD: US Department of Health and Human Services. National Institutes of Health. Available from http://nccam.nih.gov/health/whatiscam/ (Accessed 2013 January 30).
2. Lesho EP. An overview of osteopathic medicine. *Arch Fam Med* 1999;8(6):477-84.
3. UK National Health Service Careers: Complementary and alternative medicine. UK Department of Health. Available from http://www.nhscareers.nhs.uk/explore-by-career/wider-healthcare-team/careers-in-the-wider-healthcare-team/clinical-support-staff/complementary-and-alternative-medicine-(cam)/ (Accessed 2013 July 12).
4. Barnes PM, Bloom B, Nahin RL. Complementary and alternative medicine use among adults and children: United States, 2007. *Natl Health Stat Report* 2008(12):1-23.
5. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M, Kessler RC. Trends in alternative medicine use in the United States, 1990-1997: results of a follow-up national survey. *JAMA* 1998;280(18):1569-75.
6. Ernst E. Prevalence of use of complementary/alternative medicine: a systematic review. *Bull World Health Organ* 2000;78(2):252-7.
7. Nahin RL, Barnes PM, Stussman BJ, Bloom B. Costs of complementary and alternative medicine (CAM) and frequency of visits to CAM practitioners: United States, 2007. *Natl Health Stat Report* 2009(18):1-14.
8. Thomas KJ, Nicholl JP, Coleman P. Use and expenditure on complementary medicine in England: a population based survey. *Complement Ther Med* 2001;9(1):2-11.
9. Landmark Healthcare. The Landmark Report on Public Perceptions of Alternative Care. Sacramento: Landmark Healthcare, 1998.
10. Harris P, Rees R. The prevalence of complementary and alternative medicine use among the general population: a systematic review of the literature. *Complement Ther Med* 2000;8(2):88-96.
11. Acupuncture Association of Chartered Physiotherapists. AACP Background. Peterborough: AACP. Available from http://www.aacp.org.uk/index.php?option=com_content&view=article&id=55&Itemid=67 (Accessed 2013 July 12).
12. Melchart D, Weidenhammer W, Streng A, Reitmayr S, Hoppe A, Ernst E, Linde K. Prospective investigation of adverse effects of acupuncture in 97 733 patients. *Arch Intern Med* 2004;164(1):104-5.
13. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain* 2006;10(4):287-333.
14. Cochrane Central Register of Controlled Trials (CENTRAL) 2012, Issue 12, part of *The Cochrane Library*.
15. MacPherson H, Thomas K, Walters S, Fitter M. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;323(7311):486-7.
16. White A, Hayhoe S, Hart A, Ernst E. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. *BMJ* 2001;323(7311):485-6.
17. Witt CM, Pach D, Brinkhaus B, Wruck K, Tag B, Mank S, Willich SN. Safety of acupuncture: results of a prospective observational study with 229,230 patients and introduction of a medical information and consent form. *Forsch Komplementmed* 2009;16(2):91-7.
18. Langevin HM, Wayne PM, Macpherson H, Schnyer R, Milley RM, Napadow V, Lao L, Park J, Harris RE, Cohen M, et al. Paradoxes in acupuncture research: strategies for moving forward. *Evid Based Complement Alternat Med* 2011;2011:180805.
19. Barnes PM, Powell-Griner E, McFann K, Nahin RL. Complementary and alternative medicine use among adults: United States, 2002. *Adv Data* 2004(343):1-19.
20. Thomas K, Coleman P. Use of complementary or alternative medicine in a general population in Great Britain. Results from the National Omnibus survey. *J Public Health (Oxf)* 2004;26(2):152-7.
21. Ernst E, White A. The BBC survey of complementary medicine use in the UK. *Complement Ther Med* 2000;8(1):32-6.
22. Hunt KJ, Coelho HF, Wider B, Perry R, Hung SK, Terry R, Ernst E. Complementary and alternative medicine use in England: results from a national survey. *Int J Clin Pract* 2010;64(11):1496-502.
23. Posadzki P, Watson LK, Alotaibi A, Ernst E. Prevalence of use of complementary and alternative medicine (CAM) by patients/consumers in the UK: systematic review of surveys. *Clin Med* 2013;13(2):126-31.
24. Harris P. Difficulties of conducting systematic reviews of survey studies. *Complement Ther Med* 2000;8(4):288.
25. Hartel U, Volger E. [Use and acceptance of classical natural and alternative medicine in Germany--findings of a representative population-based survey]. *Forsch Komplementarmed Klass Naturheilkd* 2004;11(6):327-34.
26. Wu AP, Burke A, LeBaron S. Use of traditional medicine by immigrant Chinese patients. *Fam Med* 2007;39(3):195-200.
27. Ishizaki N, Yano T, Kawakita K. Public status and prevalence of acupuncture in Japan. *Evid Based Complement Alternat Med* 2010;7(4):493-500.
28. Chen FP, Kung YY, Chen TJ, Hwang SJ. Demographics and patterns of acupuncture use in the Chinese population: the Taiwan experience. *J Altern Complement Med* 2006;12(4):379-87.

29. Cherkin DC, Sherman KJ, Avins AL, Erro JH, Ichikawa L, Barlow WE, Delaney K, Hawkes R, Hamilton L, Pressman A, et al. A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med* 2009;169(9):858-66.
30. O'Connell NE, Wand BM, Goldacre B. Interpretive bias in acupuncture research?: A case study. *Eval Health Prof* 2009;32(4):393-409.
31. Clarke M, Horton R. Bringing it all together: Lancet-Cochrane collaborate on systematic reviews. *Lancet* 2001;357(9270):1728.
32. Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA* 2005;293(19):2362-6.
33. Montori VM, Wilczynski NL, Morgan D, Haynes RB. Systematic reviews: a cross-sectional study of location and citation counts. *BMC Med* 2003;1:2.
34. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490.
35. Ezzo J, Hadhazy V, Birch S, Lao L, Kaplan G, Hochberg M, Berman B. Acupuncture for osteoarthritis of the knee: a systematic review. *Arthritis Rheum* 2001;44(4):819-25.
36. Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. *Ann Intern Med* 2004;141(12):901-10.
37. Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, van der Windt DA, Berman BM, Bouter LM. Acupuncture for peripheral joint osteoarthritis. *Cochrane Database Syst Rev* 2010(1):CD001977.
38. Clarke M, Alderson P, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals. *JAMA* 2002;287(21):2799-801.
39. Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet* 2010;376(9734):20-1.
40. Chalmers I. Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In: Egger M, Davey SG, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 429-443.
41. Young C, Horton R. Putting clinical trials into context. *Lancet* 2005;366:107-08.
42. US National Institutes of Health. National Institutes of Health Public Access. Bethesda: US Department of Health and Human Services. National Institutes of Health. Available from http://publicaccess.nih.gov/ (Accessed 2013 July 12).
43. The Cochrane Collaboration. Global data usage. Available from http://www.thecochranelibrary.com/view/0/FreeAccess.html (Accessed 2013 July 12).
44. Alfaleh K, Anabrees J, Bassler D, Al-Kharfi T. Probiotics for prevention of necrotizing enterocolitis in preterm infants. *Cochrane Database Syst Rev* 2011;3:CD005496.
45. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327(4):248-54.
46. Dickersin K, Manheimer E. The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin Obstet Gynecol* 1998;41(2):315-31.
47. University of Maryland, School of Medicine, Cochrane CAM Field: Cochrane CAM Field journal columns. Available from http://www.compmed.umm.edu/journal_columns.aspx (Accessed 2013 July 12).
48. University of Maryland, School of Medicine, Cochrane CAM Field: Plain Language Summaries/Summaries of Findings. Available from http://www.compmed.umm.edu/cochran-plainlang.asp (Accessed 2013 July 12).
49. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;18(4):476-99.
50. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, Gunther K, Hauselmann H, Herrero-Beaumont G, Kaklamanis P, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis* 2003;62(12):1145-55.
51. Richmond J, Hunter D, Irrgang J, Jones MH, Snyder-Mackler L, Van Durme D, Rubin C, Matzkin EG, Marx RG, Levy BA, et al. American Academy of Orthopaedic Surgeons clinical practice guideline on the treatment of osteoarthritis (OA) of the knee. *J Bone Joint Surg Am* 2010;92(4):990-3.
52. The care and management of osteoarthritis in adults. Clinical guideline no. CG59. London: National Institute for Health and Clinical Excellence, 2008.
53. Low back pain: early management of persistent non-specific low back pain. Clinical guideline no. 88. London: National Institute for Health and Clinical Excellence, 2009.

54. Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147(7):492-504.
55. Chou R, Qaseem A, Snow V, Casey D, Cross JTJ, Shekelle P, Owens DK. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147(7):478-91.
56. Ammendolia C, Furlan AD, Imamura M, Irvin E, van Tulder M. Evidence-informed management of chronic low back pain with needle acupuncture. *Spine J* 2008;8(1):160-72.
57. Allen C, Richmond K. The Cochrane Collaboration: International activity within Cochrane Review Groups in the first decade of the twenty-first century. *J Evid Based Med* 2011 Feb;4(1):2-7.
58. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4(3):e78.
59. Manheimer E, Wieland S, Kimbrough E, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for Traditional Chinese Medicine therapies. *J Altern Complement Med* 2009;15(9):1001-14.
60. Wieland L*, Manheimer E*, Berman B. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane Collaboration. *Altern Ther Health Med* 2011;17(2):50-9. *Both authors contributed equally to this work.
61. Wieland LS, Manheimer E, Sampson M, Barnabas JP, Bouter LM, Cho K, Lee MS, Li X, Liu J, Moher D, et al. Bibliometric and content analysis of the Cochrane Complementary Medicine Field specialized register of controlled trials. *Syst Rev* 2013;2:51.
62. Manheimer E* and Ezzo J*. Chapter 2: Systematic reviews and CAM. In: Adams J (editor). *Researching Complementary and Alternative Medicine*. Oxford: Routledge, 2006, pgs 19-36. *Both authors contributed equally to this work.
63. Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. *J Clin Epidemiol* 2006;59(2):107-13.
64. Manheimer E, van der Windt D, Cheng K, Stafford K, Liu J, Tierney J, Lao L, Berman BM, Langenberg P, Bouter LM. The effects of acupuncture on rates of clinical pregnancy among women undergoing in vitro fertilization: a systematic review and meta-analysis. *Hum Reprod Update* 2013 Jun 27;doi:10.1093/humupd/dmt026. [Epub ahead of print]
65. Manheimer E, White A, Berman B, Forys K, Ernst E. Meta-analysis: acupuncture for low back pain. *Ann Intern Med* 2005;142(8):651-63. Erratum in: *Ann Intern Med*. 2005 Jun 7;142(11):950-1.
66. Becker L. The Cochrane Collaboration's prioritization approaches. US Cochrane Center Conference on Priority Setting for Systematic Reviews. July 10, 2008, Baltimore, USA.
67. Nasser M, Welch V, Ueffing E, Tugwell P, Crowe S, Liberati A. The proposed Cochrane agenda setting and priority setting methods group: establishing a research program. Oral presented at: 19th International Cochrane Colloquium; 2011 Oct 19-2; Madrid. p. 21. Abstract no. B4.
68. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007;147(4):224-33.

Chapter 2. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane Collaboration

Wieland LS*, Manheimer E*, Berman BM. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane Collaboration. Altern Ther Health Med 2011 Mar-Apr;17(2):50-9. *Both authors contributed equally to this work.

ABSTRACT

Over the last decade the Cochrane Collaboration has been an increasingly important source of information on complementary and alternative (CAM) therapies. From 2007 to 2008 the Cochrane CAM Field developed a Field topics list that allowed us to categorize all 396 Cochrane reviews related to CAM (as of Issue 4, 2009). This topic list is an advance in making Cochrane reviews on CAM topics accessible to the public. In this paper, we discuss challenges in developing this topic list, including developing an operational definition of CAM, deciding which reviews should be included within the CAM Field's scope, developing the structured list of CAM Field-specific topics, and determining where the reviews should be placed in the CAM Field-specific topic list. Even though aspects of our operational definition of CAM may be open to question, a standardized definition at least provides us with an objective, reproducible and systematic method for defining and classifying CAM therapies.

BACKGROUND

Over the last decade the Cochrane Collaboration has been an increasingly important source of information on complementary and alternative (CAM) therapies.[1,2] Because CAM therapies are relevant to multiple health care conditions, members of the Cochrane CAM Field felt that it might be useful to develop a convenient way for users of *The Cochrane Library* (researchers, clinicians, consumers) to identify reviews that are CAM-related, and to find reviews on specific CAM therapies. From 2007 to 2008 the Cochrane CAM Field developed a Field topics list categorizing all 396 CAM-related Cochrane reviews (as of Issue 4, 2009). This topics list, which is part of the internal Cochrane administrative database and facilitates internal Field operations and communications among Cochrane colleagues, is also an advance in making Cochrane reviews on CAM topics accessible to the public. The topic list, which is freely available online at http://www.cochrane.org/reviews/en/topics/22_reviews.html, provides links to the titles and abstracts of all Cochrane reviews on CAM therapies. In addition, the Field topic list clarifies, to internal and external audiences, the scope of the CAM Field. This paper describes the rationale for producing the Field topics list and reviews some of the obstacles we encountered in operationalizing and classifying CAM therapies in our Field topics list. We conclude with a discussion of ways in which the Field topic list may assist in identifying gaps in the Cochrane review literature on CAM.

CHALLENGES IN DEVELOPING THE CAM FIELD TOPICS LIST

We encountered three major challenges in developing the CAM Field topic list. Our first challenge was to develop an operational definition of CAM so that we knew which reviews to classify as CAM-related. Our second challenge was to apply this operational definition. That is, what methods should we use to classify individual reviews on *The Cochrane Library* as meeting or not meeting the operational definition eligibility criteria? Our third major challenge was to develop a classification tree structure to organize the CAM-related reviews into sub-categories in a list, so that individual reviews on specific therapies could be easily found.

Developing an operational definition of CAM

The set of CAM interventions is comprised of multiple therapies. During the last decade and a half, some of these therapies have been integrated into medical school curriculums and provided alongside conventional therapies in physician offices, clinics and hospitals.[3,4] CAM therapies have also been the focus of randomized controlled trials and systematic reviews, and some authorities have suggested distinguishing between unproven versus proven treatments, rather than conventional versus unconventional therapies.[5] Despite this increased mainstream openness to and acceptance of CAM, there is a shared sense among practitioners, researchers and consumers that there remains a group of therapies that are in some sense outside the mainstream medical model, and that these therapies are appropriate to group together under the CAM label. However, while some therapies would be accepted by virtually everyone as CAM (e.g., acupuncture) other therapies would be accepted as CAM by some people and not by others (e.g., vitamin supplements). The therapies that are members of the CAM group therefore vary, and this lack of an existing set of identified CAM therapies was the first challenge we encountered in developing our Field topics list.

Operational versus theoretical definition

As noted above, in order to decide whether or not specific Cochrane reviews fell within the scope of the CAM Field, we needed to develop an operational definition of CAM. In contrast to a *theoretical* definition, which characterizes the fundamental nature of a construct, an *operational* definition tests

whether a specific instance is or is not a member of the construct through a series of criteria or tests. A theoretical construct may have many different possible operational definitions. For example, in a randomized trial of depression medication, the theoretical construct of the disease being treated is 'depression', but the operational definition may be, for example, *score of 20 or more on the Beck Depression Inventory* or *answering 'yes' to the question 'are you depressed'*. Without specific operational criteria the theoretical definition is of limited practical use.

The definition of CAM has been much debated in recent years as CAM has increasingly become a focus of public and academic attention. While there have been many theoretical definitions of CAM[6,7] there are no comprehensive operational definitions of what should be considered CAM.

The need for an operational definition of CAM to construct a set of Cochrane CAM-related reviews

Because there is no accepted operational definition of CAM, we found that each time we tried to classify Cochrane reviews as CAM or not CAM we would end up with overlapping but slightly different sets of reviews. Because of this difficulty, we decided that we had to develop and implement an operational definition in order to build the CAM Field topic list. This is not to say that the operational definition we developed is the 'right' one, only that we needed to create explicit and transparent criteria in order to clarify the scope of the CAM field and build our list of CAM-related Cochrane reviews. Otherwise we would have no objective, reproducible and systematic criteria for including or excluding individual Cochrane reviews from our scope and we would continue to have the problem of including different sets of reviews as CAM each time we populated our list.

The need for an operational definition of CAM to facilitate and harmonize research

Standardization of definitions of CAM also means that different groups working independently on database development can collaborate more effectively and build on each others' work. For example, not only the Cochrane CAM Field, but also the UK National Health Service is developing a database of CAM studies; if we use the same definitions, then we can collaborate more easily, avoid duplication of effort, and promote harmonization of processes.

Standardization is also useful in promoting and comparing research in different CAM areas. For example, defining CAM operational criteria allows survey results on CAM prevalence to be comparable across time and across investigators. Depending on how CAM is defined, prevalence surveys on CAM use vary from 10% to 100%.[8] It seems like an enormous waste of research resources to not allow for the results of these different studies to be put side by side. If operational criteria are consistent, then studies can be directly compared. If operational criteria are transparent, then even if there is inconsistency between different sets of criteria there is some basis for comparison.[9]

As another example, standardized operational criteria allow examination of evidence for benefits or adverse effects of CAM. That is, often people claim that there is no evidence for CAM therapies. However, this statement is not interpretable if there is no clear agreement on what is CAM. What if one were to wish to compare the evidence base for CAM versus the evidence base for conventional medicine? If there is no clear definition of what CAM is, how could we do this? Indeed, it is misleading to even use the word CAM, if different people would define it in different ways.

Using a theoretical definition of CAM for clues to develop an operational definition

To begin to develop an operational definition, we started by looking at the theoretical definitions of CAM. The widely-accepted theoretical definition of CAM (see Box 1) was arrived at by the Office of

Alternative Medicine (OAM) expert panel at the Conference on CAM Research methodology in April 1995. (The Office of Alternative Medicine later became the US National Center for Complementary and Alternative Medicine (NCCAM).)

Box 1. Theoretical definition of CAM

> "Complementary and alternative medicine (CAM) is a broad domain of healing resources that encompasses all health systems, modalities, and practices and their accompanying theories and beliefs, other than those intrinsic to the politically dominant health system of a particular society or culture in a given historical period."

This definition is not an operational definition because it does not tell you whether or not, for example, acupuncture or relaxation therapy or omega-3 supplementation are CAM therapies. You could think about whether or not an individual therapy meets the definition above, but the definition does not provide concrete tests to tell you 'yes' or 'no' about any specific therapy. However, the theoretical definition does provide a major criterion we needed to consider in developing our operational criteria. We needed to take into account whether the therapies are founded upon the theories of disease and healing that are fundamental to the dominant health system in the culture in which the Cochrane CAM field is situated – the conventional Western medical model as it is practiced in the early twenty-first century. That is, therapies such as acupuncture or homeopathy that are based on non-allopathic theories of disease and healing are clearly CAM. One path to developing an operational definition of CAM could therefore involve identifying and listing therapies that rely upon non-allopathic models of health.

The major limitation to considering this criterion, and solely this criterion, in developing an operationalization of CAM is that the dominant medical model changes over time. As we have seen above, despite a common understanding that acupuncture is indeed 'CAM', many CAM therapies, including acupuncture, have experienced increased inclusion in Western medical contexts during the past several years. If one were to take the view that whether or not something is CAM depends on the evidence accepted by the dominant system, then defining what is CAM would require a periodic re-evaluation as the evidence changes over time, and the politically dominant health system incorporates the therapy not only in practice but also in principle. For example, what if acupuncture becomes explicable in terms of allopathic mechanisms and is therefore completely accepted within conventional medicine? What if acupuncture is proven highly effective for some indications and thus becomes an accepted treatment even though the allopathic model does not provide a plausible mechanism of action? What if the allopathic model itself evolves to include the concept of Qi? In each of these cases we would be forced to ask whether acupuncture is no longer CAM, despite its origins outside conventional medicine, and if it is no longer CAM then what was the defining moment when it lost its CAM status. Furthermore, at some historical times there may be multiple medical paradigms that coexist without any single model achieving clear dominance. For now, we considered these questions but decided to concentrate on defining the parameters of CAM from our current historical and geographical perspective. However, we felt that we needed to go beyond the non-allopathic criterion in operationalizing CAM.

Criteria considered in developing CAM operational definition

In expanding and refining our CAM operational definition, we considered several criteria. First we considered whether the historical notion of the therapy was CAM or conventional. As suggested above in the theoretical model of CAM, if the therapy was based upon the theories of a medical system outside the Western allopathic medical model, then (from the current perspective of the US and Europe, anyway) it would be labeled *alternative* medicine, or CAM. We therefore included therapies such as Chinese and Japanese traditional medicine, Ayurvedic medicine, homeopathy, chiropractic, naturopathy and Reiki, among others, as CAM, and we did so regardless of their current level of acceptance by the allopathic system.

For therapies that did not clearly originate outside the theories or beliefs of the allopathic medical system, we next considered whether the use of the therapy for a particular condition is currently considered to be a standard treatment within the dominant medical system. If something is currently a standard, accepted therapy, then it is not likely to be widely considered as CAM. Indicators that a therapy is accepted include government licensing of practitioners, coverage by health insurance, statements of approval by government agencies, and recommendation as part of a practice guideline. For example, the US Federal Drug Administration (FDA) has approved a device for administering transcranial magnetic stimulation (TMS) in treatment resistant depression (http://www.medinewsdirect.com/?p=593). TMS for depression may therefore be questionable as a CAM therapy. While following FDA guidelines would only be helpful for a limited number of therapies and would be exclusive to the US, we considered this to be one source of information [not definitive] on whether or not a specific therapy is CAM. Using the criterion of acceptability to the dominant health system meant that we would have to consider treatments and conditions together, since a treatment might be standard for one condition but alternative for another medical condition. For example, TMS is not approved by the FDA as a treatment for Parkinson Disease. As we stated above, in cases where we had decided that the therapy was CAM based upon its origins outside the allopathic system, we did not change our assessment if the therapy was accepted by insurance or practice guidelines. For example, some health insurance plans in the US and Europe cover acupuncture or chiropractic treatment, but we nevertheless considered these therapies to be CAM.

Third, for therapies that did not clearly originate outside of the allopathic system, we also considered the setting in which the therapy is delivered. Therapies that are self-care or delivered by alternative practitioners are more likely to be widely considered as CAM, while therapies that are delivered exclusively by conventionally credentialed medical personnel or exclusively within hospital settings are much less likely to be considered as CAM. Thus, nutritional therapies, including special diets and dietary supplements, are likely to be widely considered as CAM, while surgery is not considered CAM. Some therapies would be classified differently according to this consideration versus that of the previous criterion. For example, a self-delivered nutritional therapy that is supported by FDA recommendations might be considered as CAM by some persons and as non-CAM by others.

Finally, we should mention that we did not consider evidence of efficacy (or lack of evidence) as a test for identifying a CAM therapy. This is because there are many therapies that are not currently accepted as efficacious, but not all of them would be necessarily considered CAM. For example, a new synthetic chemotherapy agent would not be considered CAM, even if it has not been proven to be efficacious, while an herbal therapy for cancer would generally be considered CAM, even where it had trial evidence of efficacy. As with CAM therapies, many conventional therapies also do not have convincing evidence of benefit. An earlier assessment of reviews from the Cochrane Collaboration found that less than 25% of Cochrane reviews on conventional biomedical

interventions resulted in significant evidence of benefit.[10] Therefore, lack of proven efficacy is not an appropriate test for CAM.

With these criteria in mind, and a beginning set of CAM therapies, we looked for sources that would give us additional guidance in developing a list of specific therapies for our operational definition.

Sources considered in developing the operational definition of CAM

We examined several existing sources in developing the CAM operational definition. First, we looked at two sources within the US National Library of Medicine's PubMed database: 1) the MeSH (Medical Subject Headings) definition of complementary therapies and 2) the Complementary Medicine subset search strategy. The MeSH definition of complementary therapies is "Therapeutic practices which are not currently considered an integral part of conventional allopathic medical practice." This is a theoretical rather than an operational definition, and therefore was of limited use for our purposes. The Complementary Medicine *subset* search strategy conducts a complex search within MEDLINE, with the terms in the search strategy representing different CAM therapies.[11] It is a form of an operational definition. However, the subset search strategy is not precise as to therapies/conditions pairings, and it is not sensitive or specific enough.

We next considered conducting a Delphi survey, which is a structured series of surveys consisting of rounds of opinion collection and feedback, to arrive at an operational definition of CAM. However, we decided against it, primarily because of time and resource limitations. We did consult recent research utilizing Delphi surveys, and confirm that we did not omit any CAM therapies identified by this strategy from our own operational definition.[9]

CAM operational definition – list of therapies included

Based upon our review of the sources described above, we identified CAM therapies as therapies used in treating or preventing disease that were captured by the 70 different terms or combination of terms listed in Table 1.

Table 1. List of therapies included as CAM

Acupressure	Imagery (i.e., visualization techniques)
Acupuncture (e.g., needle acupuncture, electroacupuncture)	Light therapy† (phototherapy)
Alexander technique	Magnetic field therapy† (eg, transcranial magnetic stimulation)
Aromatherapy	Massage
Arts therapy (e.g., dance therapy, drama therapy, music therapy)	Meditation
Ayurvedic traditional medicine (Ayurveda)	Morita therapy
Balneotherapy	Moxibustion
Bee products (eg, honey, pollen, propolis, royal jelly, venom)	Naturopathy
Biofeedback	Osteopathic manipulation
Chelation therapy†	Ozone therapy†
Chinese traditional medicine	Play therapy
Chiropractic (i.e., spinal manipulation)	Prolotherapy
Color therapy (i.e., chromotherapy)	Qi gong
Craniosacral manipulation	Reflexology
Dietary supplements (non-herbal)† (e.g., vitamins, hormones, amino acids)	Reiki therapy
Diet therapy† (e.g., low fat diets, vegan diets)	Relaxation techniques
Distant healing	Snoezelen
Electric stimulation therapy† (eg, transcutaneous electrical nerve stimulation)	Speleotherapy
Electromagnetic therapy†	Spiritual healing (eg, prayer)
Eye Movement Desensitization and Reprocessing (EMDR)	Tai chi
Feldenkrais method	Therapeutic touch
Herbal supplements (eg, echinacea, garlic)	Traditional healers and healing practices (other than Chinese) (eg, Kampo, Shamanism)
Homeopathy	Tui na
Hydrotherapy	Ultrasonic therapy†
Hyperbaric oxygenation†	Yoga
Hypnosis	

†Depending upon the condition being treated, these therapies may also be standard Western allopathic treatments.

We attempted to be comprehensive in our operationalization of CAM therapies, even including multiple terms that could be considered to overlap (e.g., Chinese traditional medicine and acupuncture). However the number of individual CAM therapies within some categories (e.g., individual herbs within the herbal therapies category) was impossible to capture in a single table suitable for publication. This operationalization is also biased towards inclusion of therapies that have been the subject of randomized controlled trials, and there may be other CAM therapies of which we are unaware. This operationalization therefore cannot be considered to be exhaustive, and is subject to expansion over time.

For many of these categories, the operationalization is clear-cut and everything described with the term would be considered to be CAM. For example, any therapy described as 'homeopathy' or 'homeopathic' would be considered complementary and alternative. In some cases, however, as described above in our discussion of the criteria for CAM, the context of the therapy determines

whether the therapy would be considered CAM or not, and we needed to be specific about the therapy/condition pairing because the therapy was conventional and accepted for some conditions but complementary and alternative for others. For example, hyperbaric oxygenation is a standard treatment for carbon monoxide poisoning, but is an alternative treatment for multiple sclerosis. We have omitted the therapy/condition pairings from Table 1 for reasons of space, although we have indicated with a note those therapies that are considered conventional in some situations.

The category in which we had to be most concerned about therapy/condition pairings was dietary supplements. Our first exclusion in this category was supplements that are administered parenterally in hospital settings (e.g., intravenous magnesium for acute traumatic brain injury). Our rationale for this was that both the setting and the route of administration are so embedded in the dominant health care system that most persons would not consider the therapy to be CAM. A second category of exclusion was dietary supplementation for treatment or prevention of medically diagnosed deficiency states and disorders (e.g., iodine supplementation for preventing iodine deficiency disorders in children), and a third major exclusion was vitamin treatments used for preventing or treating disease in countries where vitamin deficiency is widespread (e.g., Vitamin A for treating measles in children in Niger). We excluded these last two categories of reviews because the therapy is provided in the context of a conventional determination of deficiency, and we therefore believe that almost all researchers, and probably most consumers, would not consider these to be examples of CAM. It was more difficult to decide whether to include as CAM dietary therapies that are accepted for prevention or treatment of specific disorders. For example, we debated whether to include the review of folic acid for neural tube defects as CAM-related. Many would not consider this treatment-condition pairing to be CAM-related because it has strong supportive evidence from randomized controlled trials which has resulted in its being integrated into the dominant healthcare system, such that folic acid supplementation is approved by the FDA for prevention of neural tube deficits. Because dietary supplements and therapies are generally self-administered and not dependent on medical professionals, however, and because we believed that users of our Topic List would expect to see folic acid reviews listed under the "Vitamins" subheading, we decided to classify the folic acid review as CAM-related. In general, we decided that we should be over-inclusive rather than under-inclusive with nutritional therapies, aside from the major exclusions detailed above.

Our last categories of exclusion from the CAM operationalization were not based on therapy/condition pairings. We decided to exclude reviews of exercise therapies, with the exception of mind body exercise (e.g., tai chi, yoga), and psychotherapy, with the exception of unconventional psychotherapies (e.g., Morita therapy).

We have posted our complete operational definition of CAM, with exclusions for each therapy indication where appropriate, on our Field website at www.compmed.umm.edu/Camdef.asp, where it can be updated when necessary and referred to by those interested in the scope of the Cochrane CAM Field.

Applying the operational definition to identify relevant Cochrane reviews

The second challenge in developing a topic list was deciding how to apply the operational definition to searching for and identifying relevant Cochrane reviews. We needed to set out these rules in order to make the selection of reviews transparent, consistent, and reproducible. In deciding how to apply the operational definition to identification of relevant reviews we had to ask ourselves a series of questions.

First, which text do you examine for each review on *The Cochrane Library* to select which reviews meet your operational definition eligibility criteria? We decided that we would search the title and

abstract text rather than read the full review. This strategy would identify nearly all relevant reviews, and all reviews for which a CAM therapy is an important focus of the review, and would be the most efficient approach.

Second, should retrieval be limited to reviews that find a trial with a CAM therapy? That is, what if a trial of the CAM therapy is searched for but no trials are found by the reviewer? We decided that we would include the review as long as the abstract explicitly mentioned a CAM therapy. So if the Selection Criteria in the abstract of a review specify that all trials using acupuncture were searched for, we would include the review as CAM whether or not any acupuncture trials were found. This would identify not only reviews in which trials had been found, but also reviews in which trials had been sought but not found, and would therefore assist in detecting gaps in the literature.

Third, should we only include reviews that explicitly search for therapies that we have defined as CAM? For example, in the review 'Wound cleansing for pressure ulcers', the SELECTION CRITERIA was as follows: "Randomised controlled trials (RCTs) comparing wound cleansing with no wound cleansing, or different wound cleansing solutions, or different cleansing techniques, were eligible for inclusion if they reported an objective measure of pressure ulcer healing." Note that this does not explicitly specify any CAM intervention. However, the Results section of the abstract includes an RCT of a saline spray containing Aloe Vera, which is an herbal therapy and therefore eligible as CAM. We decided that even if the review did not explicitly search for CAM therapies, the mention of a CAM therapy in the Results section would qualify the review for inclusion as a CAM review. In sum, the specific mention of a CAM therapy in the title or abstract is required for us to consider the review as CAM, and any mention of searching for or retrieving a trial of a CAM therapy requires us to include the review.

Deciding on the classification tree structure for the topics list

How to organize the list of therapies? Alphabetical or grouped conceptually?

The final challenge in developing a topic list was deciding on the classification tree structure for organizing the therapies. One possible organizing principle was simply to list the therapies alphabetically. An advantage of this organization method is that it is easy to find the therapy you are looking for in the alphabetical list. However, reviewing all therapies in the list is time-consuming. There may be some confusion about where to find therapies that could be identified with more than one name. For example, one would have to search under 'L' for light therapy and 'P' for phototherapy. Neither term is more correct than the other, and this could result in confusion if only one term is listed, and unacceptable length and opportunity for error if both terms are listed. Finally, an alphabetical list does not have a conceptual rationale. If therapies were organized by types instead of purely alphabetically it would be easier for the topic list user to dig down into the heading that would lead to the specific therapy of interest. In addition, within many of the therapies in an alphabetical list, you would have to list additional subcategories. For example, under dietary supplements you might want to make categories for enzymes, herbs, vitamins etc., unless you were willing to have the user of the topic list click on dietary supplements and be confronted with every single review related to every type of dietary supplement (i.e., over one hundred reviews), or list each dietary supplement individually (and under each name it might be known by) in the alphabetical list. For these reasons, we decided that an alphabetical approach alone would be insufficient to organize the topic list.

Classification tree structures present in existing electronic resources of CAM

At this point, we decided to take a step back and look at borrowing from existing structures to develop our own classification tree. We examined the structures of the following three systems that already exist online in CAM: the UK NHS Evidence -- CAM specialist collection (NHSE), the US National Library of Medicine (NLM) and its Medical Subject Headings (or MeSH), and the US National Library of Medicine Center for Complementary and Alternative Medicine (NCCAM).

We began by looking at the NHSE, which is an extensive online library of evidence, education, and patient information on CAM. The portal for the NHSE lists the following categories of CAM therapies alphabetically: "acupuncture, aromatherapy, chiropractic, dietary and nutritional therapies, herbal medicine, homeopathy, hypnosis, massage, meditation, osteopathy, reflexology, yoga, other therapies or medical systems". Clicking on 'herbal medicine' or 'dietary and nutritional therapies' brings up groups of herbs listed alphabetically. Clicking on "other therapies or medical systems" brings up a list of 15 additional therapies. We decided to use all of the major categories from the NHSE as major categories in our CAM Field scheme. Therefore the two schemes would be compatible, leaving open the possibility of future collaboration. However, the scheme appears to rely upon an alphabetical listing of important therapies. We decided to look further and see whether we could construct a tree structure based on a more conceptual rationale and without relying on putting multiple therapies under the category of "other", thinking that a more specific structure might be easier to navigate.

We then looked at NLM's MeSH heading structure for the MeSH term Complementary Therapies. This list of MeSH headings is also organized alphabetically. There are many differences between the Complementary Therapies (MeSH) tree and the NHSE structure described earlier. As one example, in the MeSH headings acupressure is listed as its own major heading while in the NHS headings acupressure was listed under the category 'other'. The MeSH placement of acupressure might make it easier to locate quickly. The MeSH tree does not include an 'other' category. However, there are some anomalies in the MeSH structure. For example, some of the major headings are of relatively minor therapies. Anthroposophy, which derives from the philosophy of Rudolph Steiner, has only 134 citations in all of Medline/PubMed and probably should not be a major heading. Also, some of the organization of the MeSH list is not intuitively obvious or logical. For example, the alternative medical systems of Homeopathy and Naturopathy are major headings but non-western alternative medical systems (e.g., Ayurvedic Medicine) are grouped together as "traditional medicine". We thought that perhaps we could build on aspects of the MeSH structure but look elsewhere for an overarching organizational scheme and thus avoid some of the limitations we saw here.

The last structure we looked at was that developed at the US National Institute of Health's NCCAM. NCCAM groups CAM practices into four domains, or types of therapies, recognizing there can be some overlap. In addition, NCCAM has a category of CAM whole medical systems, which cut across all domains.

The NCCAM categories of therapy are as follows[12]:

- Mind-Body Medicine, which uses a variety of techniques to enhance the mind's capacity to affect bodily function and symptoms.
- Natural Product Based Therapies, which use substances found in nature to promote health.
- Manipulative and Body-Based Practices, which are based on manipulation and/or movement of parts of the body.
- Energy Medicine, which involves the use of energy fields, either the unconventional use of electromagetic fields, or the manipulation of energy fields that purportedly surround and penetrate the human body.

- Whole Medical Systems, which are complete systems of theory and practice outside the conventional allopathic model.

A great appeal of this way of organizing CAM is that there are overarching principles by which the therapies can be organized. A possible problem with this system is that someone who does not understand the language and principles of CAM might have some difficulty in knowing what domains a specific therapy would fall under. In some cases, even someone who is familiar with CAM would have several valid options for where a therapy may be classified.

Classification suggestions from the CAM literature

Finally, we reviewed the CAM literature to identify additional proposals for how to organize and classify CAM therapies. Suggestions for organizational schema included epistemological perspective and type of healing method,[13] a taxonomy of five sectors of CAM,[14] four paradigms of health and disease,[15] and modes of therapeutic action.[16] While each of these classification schema had merit and each provided valuable insights into the parameters of CAM and the underpinnings of CAM practices, we did not feel that any one of them was sufficiently well-known or widely used to adopt for the Cochrane Field topics list.

Final classification scheme

We decided to follow the NCCAM model because of its logical domain classification system, and its prevalence online and in the literature. We found additional reputable online resources on CAM (e.g., the Mayo clinic website at http://www.mayoclinic.com/health/alternative-medicine/PN00001/NSECTIONGROUP=2), books[17] and publications[18] that discussed CAM therapies by using the NCCAM framework (sometimes with minor modifications). Because the NCCAM is a large funding organization and source of government information to researchers and consumers of CAM, we felt that researchers and consumers, especially in the US, would be most familiar with the NCCAM classification scheme and therefore this organization of topics would be the most accessible.

After deciding on this classification system, we placed each of the CAM therapies into one of the five NCCAM domains. Although there is some overlap between the domains, we put each therapy into only one place in the classification tree. For example, acupuncture could belong to Alternative Medical Systems because it is part of Chinese Traditional Medicine, or it could belong to Energy Therapies because it relies upon the principles of energy flow. Based on consensus among the authors, we chose to put it in only one of those places (i.e., we decided to classify it as an Energy Therapy). Therapies were put in only one place in the tree for practical reasons, as the Cochrane topic list software is not designed for setting up classifications and links in which therapies can be listed once and appear under multiple headings. Manually placing therapies in multiple places in the topic list would require additional resources, and present opportunities for error that would require time and energy to guard against. As mentioned above, it might be difficult for some people to guess that acupuncture is under Energy Therapies and not Alternative Medical Systems. Some topic list users might even guess that acupuncture is under Mind-Body Medicine or Manipulative and Body-Based Practices. But with only five broad categories to look under, it should become quickly apparent where each individual therapy is classified. When we were unsure which area to classify a therapy under, we looked first at whether NCCAM had explicitly included a therapy under a specific domain. If this was unclear, we looked at MeSH headings and NHS classifications for options.

Table 2 shows the fully expanded CAM Field topic list, which displays all major topics and sub-topics, and the number of Cochrane reviews associated with each topic. To organize the topics, we

relied upon some of the hierarchies we had observed in the NHSE and the National Library of Medicine MeSH headings. We tried to strike a balance between being overly broad (which would require the topic list user to search through too many reviews within a single heading to find the topic of interest) and being overly specific (which would increase the possibility of misclassification of reviews and require duplicate entry of some reviews). For example, under Natural Product Based Therapies there is a heading for Nutrition therapy, and under that is Dietary Supplements. The dietary supplements have been categorized into types of supplements. Vitamins is one of these subheadings. We did not create further categories of Vitamin A, Vitamin B, Vitamin C etc., because if we had done so, we would have had to enter the review "Vitamin supplementation for preventing miscarriage", which examined the effects of several different vitamin supplements, under multiple vitamin categories, or created the category 'Other vitamins' or 'Multiple vitamins'. In contrast, most (though not all) herbal medicines are studied individually in Cochrane reviews, and therefore we created a discrete category for each commonly used herb for which a Cochrane review exists, and an 'Other plants or plant extracts' category for Cochrane reviews examining other herbs. Finally, systematic reviews that encompass multiple CAM therapies are classified under the subtopic of each therapy reviewed. For example, the review 'Herbal medicine for low back pain' reviews trials of devil's claw, white willow bark, and cayenne, and therefore this review was classified under all three of these herbal medicine sub-topics. We hope our tree structure succeeds in striking the balance between breadth and specificity, and that it is comprehensive and user-friendly.

Table 2. Cochrane CAM-related reviews organized by subtopics*†

▼Alternative Medical Systems (107)
- ► Ayurvedic Medicine (4)
- ▼ Chinese Traditional Medicine (87)
  - ► Chinese herbal drugs (87)
- ►Homeopathy (12)
- ►Japanese traditional medicine(1)
- ►Naturopathy (1)
- ►Tibetan traditional medicine (2)

▼Natural Product Based Therapies (354)
- ►Chelation therapy (1)
- ►Hydrotherapy (3)
- ▼Nutrition therapy (280)
  - ▼Diet therapy (22)
    - ►Calorie control or calorie restriction (2)
    - ►Carbohydrate-restricted diet (1)
    - ►Casein-free diets (1)
    - ►Fat-restricted diet (2)
    - ►Gluten-free diet (1)
    - ►High-fiber diet (2)
    - ►Low glycemic-index diet (4)
    - ►Protein-restricted diet (3)
    - ►Sodium-restricted diet (3)
    - ►Vegetarian or vegan diet (1)
    - ►Other diet therapies (2)
  - ▼Dietary supplements (258)
    - ►Amino acids (20)
    - ►Enzymes and coenzymes (7)
    - ►Other plants or plant extracts (29)
  - ►Prolotherapy (1)
  - ►Speleotherapy (1)
  - ►Topical therapies (5)
  - ▼Unconventional synthetic drugs (1)
    - ►Laetrile
    - ►Procaine (1)

▼Energy Therapies (154)
- ▼Acupuncture therapy (87)
  - ►Acupressure (5)
  - ►Acupuncture (69)
  - ►Electroacupuncture (6)
  - ►Laser acupuncture (6)
  - ►Moxibustion (1)
- ▼Breathing exercises (0)
  - Qi gong (0)
- ►Distant healing (1)
- ►Electric stimulation therapy (32)
- ►Magnetic therapy (11)
- ►Phototherapy (7)
- ►Reiki therapy (2)
- ►Therapeutic touch (3)
- ►Ultrasonic therapy (11)

▼Manipulative and Body-Based Methods (21)
- ►Alexander Technique (1)
- ►Chiropractic Manipulation/Spinal Manipulation (8)

- ►Fats (26)
- ►Hormones (10)
- ► Minerals (56)
- ►Probiotics (23)
- ►Vitamins (76)
- ►Other supplements (39)

►Oxygen therapy (5)
►Ozone therapy (1)
▼Herbal Medicine (56)

- ►African prune (1)
- ►Artichoke leaf (1)
- ►Cayenne (1)
- ►Cranberry (2)
- ►Devil's claw (1)
- ►Echinacea (1)
- ►Feverfew (1)
- ►Garlic (5)
- ►Ginkgo biloba (6)
- ►Horse chestnut (1)
- ►Kava (1)
- ►Milk thistle (1)
- ►Passiflora (1)
- ►Saw palmetto (1)
- ►St. John's wort (1)
- ►Valerian (1)
- ►White willow (1)

Craniosacral Massage (0)
Feldenkrais Method (0)
►Massage (10)
Osteopathic Manipulation (0)
►Reflexology (2)

▼Mind-Body Interventions (54)

- ►Biofeedback (3)
- ►Hypnosis (7)
- Imagery (0)
- ►Meditation (2)
- ►Play therapy (1)
- ►Relaxation techniques (7)
- ▼Sensory art therapies (24)
  - ►Aromatherapy (5)
  - ►Art therapy (1)
  - Color therapy (0)
  - ►Dance therapy (2)
  - ►Drama therapy (1)
  - ►Music therapy (14)
  - ►Other sensory therapies (1)
- ►Tai Chi (5)
- ▼Unconventional psychotherapies (1)
  - ►Morita therapy (1)
- ►Yoga (4)

*Totals include reviews in progress (protocols) and withdrawn reviews, as well as current reviews.
†Systematic reviews that encompass multiple CAM therapies (e.g., 'Complementary and alternative therapies for pain management in labour') are classified under the subtopic of each therapy reviewed

REMAINING CHALLENGES AND OPPORTUNITIES

Over-inclusiveness

As described above, we have included some nutrition and vitamin reviews (e.g., folic acid for neural tube defects) even though some people may consider these not to be CAM-related. Ideally we would have some way to mark such reviews, within the topic list, as possibly non-CAM, so that topic list users would be alerted to possible over-inclusiveness. Currently, however, this is not an option of the Cochrane topic list software. We hope that our operational definition, available on our website, will alert users to our desire to be over-inclusive rather than under-inclusive. We are open to revising our policies about nutrition and dietary supplement reviews in the future.

Placing therapies in more than one place in the topic list

We recognize that having therapies appear in only one place in the topics list might cause slight difficulty for users when they first approach the Field topic list and may have to look in more than one place before they find the therapy they are looking for. For example, under the heading 'Chinese traditional medicine', the only sub-topic listed is 'Chinese herbal drugs', but users may also expect to find reviews of acupuncture and tai chi listed under this heading. The addition of a scope note under 'Chinese traditional medicine' to indicate the location in the classification tree for acupuncture (i.e., under Energy therapies) and tai chi (i.e., under 'Mind-body interventions') would be helpful to include, once the Cochrane topic list software makes this possible. If the Cochrane topic list software could be automated to place a set of reviews in multiple places in the topic list, without reviews having to be manually added to each of the multiple locations, our concerns about errors and resource utilization would be eased, and we could easily see placing therapies under

multiple categories in the topic list. For the present, we have decided to keep therapies in single locations, but concerns about usability make us open to revisiting the issue of placing individual therapies in multiple locations in the Field topic list in the future, even in the absence of changes to the topic list software.

Defining subheadings

We continue to debate how narrowly to define topic list subheadings. For example, we are unsure whether we should have subheadings for widely used herbal therapies even when there are no Cochrane reviews for the therapy. It might be frustrating for people to go to a category (e.g., a particular herb for which there are no reviews) and find that it is empty of reviews. If we do include empty therapy headings, how should we decide on which therapies are important to include in the list even without reviews? As noted above, for herbal medicine, we currently include only categories where there is at least one review. In other cases (e.g., Qi gong), we have listed categories that are empty because they are important classes of therapies rather than individual therapeutic agents. We tried to strike a balance, showing that we are aware of important therapies even in the absence of Cochrane reviews, and yet omitting multiple empty headings for individual therapies within other categories of therapy. However, we are willing to revisit this decision in the future, depending upon user feedback.

Using the Field topic list to identify gaps in Cochrane review evidence

A standardized operational definition of CAM and the development of a topic list of CAM-related Cochrane reviews present us with an opportunity to identify areas where Cochrane reviews are needed. That is, by comparing the lists of current Cochrane CAM reviews against the contents of the Cochrane CAM Field register of randomized trials we can spot condition/treatment pairings for which randomized trials have been published but no Cochrane review has yet been completed. As one example, there are currently only four Cochrane reviews on yoga in the topics list, but a search of the Cochrane register of randomized trials identifies 230 yoga controlled trials. We are beginning a project to identify therapies with randomized controlled trials but no reviews, and hope to develop a system that may be used to contribute to prioritizing future Cochrane reviews.

CONCLUSIONS

We do not believe that the Cochrane CAM Field operational definition of CAM is definitive. Indeed, we question whether it is possible to arrive upon a definitive set of therapies that are universally agreed upon as CAM. We suspect that there will never be universal agreement upon CAM aside from a core set of therapies, and that even this agreement will be susceptible to change over time. However, we are satisfied that our operational definition is transparent, and we have posted our operational criteria on the CAM Field website to maintain this transparency.

We are also satisfied with our methods for identifying CAM-related Cochrane reviews. Like our operationalization, it is a transparent method that tends to err on the side of over-inclusiveness rather than under-inclusiveness. We would rather provide information that not all persons consider to be CAM-related than omit information that a consumer may be seeking from our Field.

We are also satisfied with our identification of a popular and usable classification scheme, although we would like additional flexibility in placing reviews in multiple places and including scope notes as we feel that would further improve the usability of the topic list. Finally, we look forward to using the topic list together with our Field register of controlled trials to identify gaps in Cochrane reviews. We believe that the development and publishing of the Cochrane CAM Field

topic list will both contribute to dissemination of the systematic review evidence on CAM and lead to identifying opportunities for advancing that evidence base.

REFERENCES

1. Ezzo J, Wright K, Hadhazy V, Bahr-Robertson M, Mac Beckner W, Covington M, Berman B. Use of the Cochrane electronic library in complementary and alternative medicine courses in medical schools: is the giant lost in cyberspace? *J Altern Complement Med* 2002;8(5):681-6.
2. Manheimer E, Wieland S, Kimbrough E, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for Traditional Chinese Medicine therapies. *J Altern Complement Med* 2009;15(9):1001-14.
3. Astin JA, Marie A, Pelletier KR, Hansen E, Haskell WL. A review of the incorporation of complementary and alternative medicine by mainstream physicians. *Arch Intern Med* 1998;158(21):2303-10.
4. Pelletier KR, Marie A, Krasner M, Haskell WL. Current trends in the integration and reimbursement of complementary and alternative medicine by managed care, insurance carriers, and hospital providers. *Am J Health Promot* 1997;12(2):112-22.
5. Fontanarosa PB, Lundberg GD. Alternative medicine meets science. *JAMA* 1998;280(18):1618-9.
6. Barrett B, Marchand L, Scheder J, Plane MB, Maberry R, Appelbaum D, Rakel D, Rabago D. Themes of holism, empowerment, access, and legitimacy define complementary, alternative, and integrative medicine in relation to conventional biomedicine. *J Altern Complement Med* 2003;9(6):937-47.
7. Caspi O, Sechrest L, Pitluk HC, Marshall CL, Bell IR, Nichter M. On the definition of complementary, alternative, and integrative medicine: societal mega-stereotypes vs. the patients' perspectives. *Altern Ther Health Med* 2003;9(6):58-62.
8. Manheimer E, Anderson BJ, Stein MD. Use and assessment of complementary and alternative therapies by intravenous drug users. *Am J Drug Alcohol Abuse* 2003;29(2):401-13.
9. Lachance LL, Hawthorne V, Brien S, Hyland ME, Lewith GT, Verhoef MJ, Warber S, Zick S. Delphi-derived development of a common core for measuring complementary and alternative medicine prevalence. *J Altern Complement Med* 2009;15(5):489-94.
10. Ezzo J, Bausell B, Moerman DE, Berman B, Hadhazy V. Reviewing the reviews. How strong is the evidence? How clear are the conclusions? *Int J Technol Assess Health Care* 2001;17(4):457-66.
11. Nahim AM. Complementary medicine -- new PubMed subset. *NLM Technical Bulletin* 2001;318:e7.
12. National Center for Complementary and Alternative Medicine. What is complementary and alternative medicine? Bethesda, MD: US Department of Health and Human Services. National Institutes of Health. Available from http://nccam.nih.gov/health/whatiscam/ (Accessed 2013 January 30).
13. Engebretson J. A heterodox model of healing. *Altern Ther Health Med* 1998;4(2):37-43.
14. Kaptchuk TJ, Eisenberg DM. Varieties of healing. 2: a taxonomy of unconventional healing practices. *Ann Intern Med* 2001;135(3):196-204.
15. Tataryn DJ. Paradigms of health and disease: a framework for classifying and understanding complementary and alternative medicine. *J Altern Complement Med* 2002;8(6):877-92.
16. Jones CH. The spectrum of therapeutic influences and integrative health care: classifying health care practices by mode of therapeutic action. *J Altern Complement Med* 2005;11(5):937-44.
17. Myers CD, Bergman J, Zeltzer LK. In: Kreitler S, Ben Arush MW, editors. *Complementary and alternative medicine use in children with cancer*: John Wiley & Sons, Ltd, 2004.
18. Nathanson I, Sandler E, Ramirez-Garnica G, Wiltrout SA. Factors influencing complementary and alternative medicine use in a multisite pediatric oncology practice. *J Pediatr Hematol Oncol* 2007;29(10):705-8.

Chapter 3. Bibliometric analysis of the Cochrane Complementary Medicine Field specialized register of controlled trials

Wieland LS, Manheimer E, Sampson M, Barnabas JP, Bouter LM, Cho K, Lee MS, Li X, Liu J, Moher D, Okabe T, Pienaar ED, Shin BC, Tharyan P, Tsutani K, van der Windt DA, Berman BM. Bibliometric analysis of the Cochrane Complementary Medicine Field specialized register of controlled trials. Syst Rev 2013 July 4;2:51.

ABSTRACT

BACKGROUND: The identification of eligible controlled trials for systematic reviews of complementary and alternative medicine (CAM) interventions can be difficult. To increase access to these difficult to locate trials, the Cochrane Collaboration Complementary Medicine Field (CAM Field) has established a specialized register of citations of CAM controlled trials. The objective of this study is to describe the sources and characteristics of citations included in the CAM Field specialized register.

METHODS: Between 2006 and 2011, regular searches for citations of CAM trials in MEDLINE and the Cochrane Central Register of Controlled Trials (CENTRAL) were supplemented with contributions of controlled trial citations from international collaborators. The specialized register was 'frozen' for analysis in 2011 and frequencies were calculated for publication date, language, journal, presence in MEDLINE, type of intervention, and type of medical condition.

RESULTS: The CAM Field specialized register increased in size from under 5,000 controlled trial citations in 2006 to 44,840 citations in 2011. Most citations (60%) were from 2000 or later, and the majority (71%) were reported in English; the next most common language was Chinese (23%). The journals with the greatest number of citations were CAM journals published in Chinese, and non-CAM nutrition journals published in English. More than one-third of register citations (36%) were not indexed in MEDLINE. The most common CAM intervention type in the register was non-vitamin, non-mineral dietary supplements (e.g., glucosamine, fish oil) (34%), followed by Chinese herbal medicines (e.g., *Astragalus membranaceus*, *Schisandra chinensis*) (27%).

CONCLUSIONS: The availability of the CAM Field specialized register presents both opportunities and challenges for CAM systematic reviewers. While the register provides access to thousands of difficult to locate trial citations, many of these trials are of low quality and may overestimate treatment effects. When including these trials in systematic reviews, adequate analysis of their risk of bias is of utmost importance.

The online version of the thesis Table of Contents includes links to the full-text of the following related material:

- Supplementary appendices;
- Preliminary work presented as an abstract at the 2009 Cochrane Colloquium:
    - Wieland LS, Manheimer E. Issues in improving the quality and increasing the size of the Cochrane Complementary and Alternative Medicine (CAM) Field specialized register. In: 17th International Cochrane Colloquium; 2009 Oct 11-14; Singapore. p. 47-48. Abstract no. P56.

INTRODUCTION

Complete identification of eligible controlled trials is an essential step in conducting a systematic review, and finding and collecting citations of controlled trials has been aims of the Cochrane Collaboration from its inception 20 years ago. As part of this mission, the Collaboration developed the Cochrane Central Register of Controlled Trials (CENTRAL), a searchable database of citations of controlled trials.[1] The Collaboration has agreements with the publishers of MEDLINE and EMBASE, ensuring that all citations of controlled trials from those databases are republished in CENTRAL. Cochrane entities (e.g., Cochrane Review Groups) also regularly submit citations of controlled trials to CENTRAL, ensuring that CENTRAL contains trial citations from multiple sources, including not only MEDLINE and EMBASE, but also regional and subject-specific databases, as well as trial citations not included in databases. Because of the Collaboration's extensive efforts at identification of trials from a range of sources, CENTRAL is considered to be the most comprehensive source of citations of controlled trials for inclusion in systematic reviews.[1]

The complete identification of eligible controlled trials can be particularly challenging for systematic reviews of CAM interventions. The disadvantage of relying exclusively upon sources such as MEDLINE for trial identification is illustrated by research conducted by Egger and colleagues. Egger and colleagues[2] analyzed the characteristics of 1,635 controlled trials included in a group of n=159 systematic reviews, including both conventional and CAM-related reviews. They found that in CAM-related systematic reviews, the proportion of non-MEDLINE-indexed trials (41%) was approximately twice that proportion seen in conventional medicine systematic reviews (21%). If the systematic review authors had searched only MEDLINE for trials, they would have missed many trials, possibly including some important trials, and the proportion missing from CAM reviews would have been double the proportion missing from conventional medicine reviews. Earlier research using various 'gold standard sets' of known trials for specific CAM interventions found that the percentage of known trials included in MEDLINE was 58% for acupuncture trials,[3] 31% for ginkgo trials,[4] and 17% for homeopathy trials.[4] Ensuring that CENTRAL contains both MEDLINE and non-MEDLINE citations of CAM-related controlled trials is therefore important for the unbiased conduct of Cochrane systematic reviews of CAM interventions.

The CAM Field maintains a specialized register of citations of controlled trials of CAM interventions, which is a 'sub register' of CENTRAL. In 1998, a bibliometric analysis of the CAM Field specialized register described it as containing 3,774 controlled trials.[5] In 2006, the CAM Field began an active program to improve the scope and size of the CAM Field specialized register. CAM Field staff and international partners in this endeavor performed searches of bibliographic databases and paper journals, and CAM Field staff performed extensive quality checks and de-duplication of all identified citations. As a result of these efforts, the CAM Field register contained a total of 43,310 CAM Field specialized register citations as of Issue 1, 2012, of *The Cochrane Library*. CAM Field specialized register citations can be retrieved from the Cochrane Central Register of Controlled Trials (CENTRAL) database of *The Cochrane Library* by searching for the tag 'SR-COMPMED' in all text. The objective of this study is to describe the sources and characteristics of the trial citations included in the CAM Field specialized register.

METHODS

In 2011 we drew up a detailed protocol (Supplementary Appendix 1) that described the eligibility criteria for including citations and the methods we had used to build the register. The protocol also pre-specified the methods we planned to use in examining the characteristics of the citations in the register. A summary of these methods, as well as the details of contributions of trial citations by international collaborators, is presented below.

Eligibility criteria

All citations in the CAM Field specialized register are required to meet the following two inclusion criteria: (1) they must be reports of controlled trials, and (2) they must be CAM-related.

We considered controlled trials to be studies meeting the inclusion criteria for CENTRAL that were formulated and agreed upon in November 1992 and are published in Chapter 6.3 of the Cochrane Handbook.[6] We considered trials to be CAM-related if they described interventions that are outside the practices and theories of disease and healing that are intrinsic to the conventional Western medical model.[7] To retrieve citations of CAM trials from MEDLINE and CENTRAL, we relied upon the CAM on PubMed search strategy, which was jointly developed by the US National Library of Medicine and the US National Institutes of Health, National Center for Complementary and Alternative Medicine and introduced in PubMed in 2001.[8,9] In cases where some uses of the intervention are accepted within conventional Western medicine and others are not (e.g., vitamin supplementation), the CAM on PubMed search strategy generally does not distinguish between conventional and unconventional uses of an intervention. Therefore, for classifying interventions as conventional or CAM, we followed the same operational criteria we had previously developed for classifying Cochrane systematic reviews as conventional or CAM.[10] Some of the major decisions about the scope of CAM were as follows: we excluded vitamins and other supplements that are administered parenterally in hospital settings, we excluded dietary supplementation for treatment or prevention of medically diagnosed deficiency states (e.g., iron supplementation for preventing or treating iron deficiency), and we excluded vitamin supplements for preventing or treating disease in countries where vitamin deficiency is widespread (e.g., vitamin A for treating measles in children in Niger). We included vitamins for other conditions, even vitamins that are accepted for the prevention or treatment of specific disorders (e.g., folic acid for preventing neural tube defects). In general, we decided that we should be over-inclusive rather than under-inclusive with vitamin therapies, aside from the three major exclusions detailed above, and therefore some of the vitamin trials in the database would not be accepted as CAM by most people. Finally, we excluded exercise interventions with the exception of mind-body exercise (e.g., yoga), and we excluded conventional psychotherapies. A full description of the CAM Field operational definition of CAM has been published previously (Supplementary Appendix 2).[10]

Methods for building the register of trials

We began the expansion of the CAM Field specialized register of trials by building upon the reference management database of nearly 5,000 CAM controlled trial citations developed during the 1990s and early 2000s by Vickers and colleagues.[5] In 2006, we began regular searches of MEDLINE in PubMed using the CAM on PubMed search strategy. In 2008, an information specialist translated the CAM on PubMed search strategy into a format for use in CENTRAL, and we replaced searches of PubMed with regular searches of CENTRAL. The rationale for replacing searches of PubMed with searches of CENTRAL is that CENTRAL includes not only controlled trial citations from MEDLINE, but also controlled trial citations from multiple other sources. These other citations, which may be from other databases or from difficult to locate sources such as trial proceedings, are identified by Cochrane contributors around the world and contributed to CENTRAL. All Cochrane groups then search CENTRAL in order to identify relevant citations that others have contributed. We began by searching CENTRAL from inception and then searched newly added citations in each subsequent issue of CENTRAL.

An important subset of CAM is traditional medicine, defined by the World Health Organization as "the sum total of knowledge, skills and practices based on the theories, beliefs and experiences indigenous to different cultures that are used to maintain health, as well as to prevent, diagnose, improve or treat physical and mental illnesses. Traditional medicine that has been adopted by other

populations (outside its indigenous culture) is often termed alternative or complementary medicine."[11] Different countries often have their own forms of traditional medicine (e.g., traditional Chinese medicine, traditional Korean medicine). Because we did not have access to the traditional medicine trial reports that are published in regional or national databases and journals, in 2008 we began contacting Cochrane colleagues and contributors to solicit the contributions of citations of traditional medicine trials published in their regions. These efforts are described below. Searches of PubMed, and then of CENTRAL, were thus complemented with searches of bibliographic databases and journals conducted by several international groups, who contributed citations of trial reports to the CAM Field for inclusion in the CAM Field register. Citations provided by contributing organizations were not restricted by publication year. As described below, two of these contributing groups also provided PDFs of the full text publications for all identified citations. Collaborators are listed below in order of numbers of citations submitted to the CAM Field for the specialized register.

Traditional Chinese Medicine (TCM) trials identified by Chinese collaborators
Beginning in 2008, staff at the Center for Evidence-Based Medicine of the Beijing University of Chinese Medicine, under the direction of Jianping Liu, searched both electronic databases and Chinese journals to identify citations of controlled trials of TCM interventions. The journal titles, article titles, and abstracts (if available) of all identified citations were translated into English, entered into a reference management database with added topic keywords, and submitted to the CAM Field for inclusion in the specialized register. The full text report was also submitted to the CAM Field for each citation, in a PDF format.

Trials from CAM-specific databases identified by Canadian collaborators
In 2008, information specialists under the direction of David Moher of the Ottawa Hospital Research Institute undertook a project to search several specialized databases for difficult to locate controlled trials of CAM interventions, and these searches were replicated in 2010.[12] All identified citations were imported into a reference management database, information about the source database and the type of CAM intervention was included for each citation, and the database was submitted to the CAM Field for inclusion in the CAM Field specialized register.

Traditional medicine trials identified by Korean collaborators
In 2010, researchers at the Korea Institute of Oriental Medicine, under the direction of Myeong Soo Lee, searched both electronic databases and journals to identify citations of controlled trials of traditional medicine interventions conducted in Korea and primarily published in non-MEDLINE journals. Initial searches were focused on identifying trials of acupuncture[13] and ginseng, and were then expanded to include all other CAM interventions. The journal titles, article titles, and abstracts (if available) of all identified citations were translated into English and entered into a reference management database with topic keywords, and the citations were submitted to the CAM Field. For each citation, the full text publication was also submitted to the CAM Field in a PDF format.

Kampo trials identified by Japanese collaborators
Kampo is the Japanese adaptation of traditional Chinese medicine. While Kampo uses most of the interventions of Chinese medicine, including acupuncture and moxibustion, its primary focus is on the study and evaluation of traditional herbal medicines. In 2001, the Japan Society for Oriental Medicine undertook a project to collect controlled trial evidence on Kampo interventions through searches of both electronic databases and journals.[14] As randomized controlled trials of Kampo interventions are identified, structured abstracts are prepared for each trial, and the citations and

structured abstracts are published online in English. In March 2011, one of the leaders of this initiative, Kiichiro Tsutani of the University of Tokyo, provided permission for the CAM Field staff to incorporate the citations associated with these Kampo trials into the CAM Field specialized register together with links to the online structured abstracts.

Ayurveda and other CAM-related trials identified by Indian collaborators
The South Asian Database of Controlled Clinical Trials (SADCCT) is an online database of citations of controlled trials that have been conducted in countries for which the South Asian Cochrane Network & Centre is the reference Cochrane Centre, including Afghanistan, Bangladesh, Bhutan, India, the Maldives, Nepal, Pakistan, and Sri Lanka. The SADCCT was developed by searching South Asian journals and conference proceedings for all controlled trials.[15] In 2011, staff at the South Asian Cochrane Network & Centre, under the direction of Prathap Tharyan, identified and forwarded citations of South Asian trials of Ayurveda, Unani, Siddha, and other CAM interventions included in the SADCCT, and in 2012 staff at the CAM Field identified additional trial citations from the online SADCCT.

CAM trials identified by African collaborators
The African Trials Register is a database of citations of controlled trials conducted in Africa. It has been developed at the South African Cochrane Centre by the Cochrane Centre staff searching African journals and electronic bibliographic databases.[16,17] In 2011, citations of CAM-related trials included in the African Trials Register were identified by Elizabeth Pienaar, information specialist at the South African Cochrane Centre, and a database of citations was forwarded to the CAM Field.

Methods for examining the contents of the register of trials
In August 2011, we suspended additions of new search results to the CAM Field specialized register and began an intensive program of cleaning and updating the register in preparation for analysis, focusing on detection and removal of any non-CAM or non-controlled trial citations, deduplication of register citations, identification of whether each register citation was present or absent in MEDLINE, and standardization of journal names (for details of procedures used, see Supplementary Appendix 1). Because EMBASE is a second major database from which controlled trials are automatically downloaded to CENTRAL, we also wished to characterize the EMBASE coverage of CAM Field register citations. However, register citations do not contain EMBASE identifiers, and we did not have the resources to comprehensively check all citations for presence in EMBASE. We therefore estimated the proportion of register citations present in EMBASE by taking a random sample of 200 register records and searching EMBASE for each citation in the sample. We used the same random sample to estimate the overlap in coverage between MEDLINE and EMBASE, and the proportion of register citations not present in either MEDLINE or EMBASE.

We sorted journal titles by frequency and the 100 journals associated with the greatest number of trial citations were classified as either CAM or conventional in focus, using the classification method described in the protocol (see Supplementary Appendix 1). For each of the 25 CAM and 25 conventional journals with the greatest number of trial citations, we determined the language of publication and whether the journal was indexed in MEDLINE. In addition to characterizing the journals with the greatest number of citations, we also examined the distribution of citations across all journals in the register, to determine to what extent citations are scattered across journals overall.

To characterize register citations by CAM intervention, we chose 21 different types of CAM interventions within five broad categories. We based the five broad categories and the 21 intervention types within categories upon the CAM Field topics list for Cochrane reviews of CAM interventions[10] as well as other classifications of CAM interventions (e.g., the classifications of CAM

interventions used in the 2007 NHIS survey of use of CAM in the United States[18]). We then developed subject searches for each of the 21 types of CAM interventions by parsing the 2006 translation of the CAM on PubMed search strategy into the CAM intervention topic areas. We also consulted additional sources to identify any supplementary terms and to help understand and delineate between the CAM intervention topic areas. Searches were run and tested in MEDLINE, using each relevant term. An information specialist developed the MEDLINE version of each search strategy, which was then peer reviewed by an independent information specialist using the PRESS standard.[19] Searches were then adapted so that they could be used to search the reference management database containing the register. MEDLINE Medical Subject Heading (MeSH) terms and free text terms were sought in all database fields. The search strategies for identifying each of the 21 different types of CAM interventions are included as an additional file, so that these strategies will be publicly available to CAM systematic reviewers developing searches for any of these classes of therapies (see Supplementary Appendix 3).

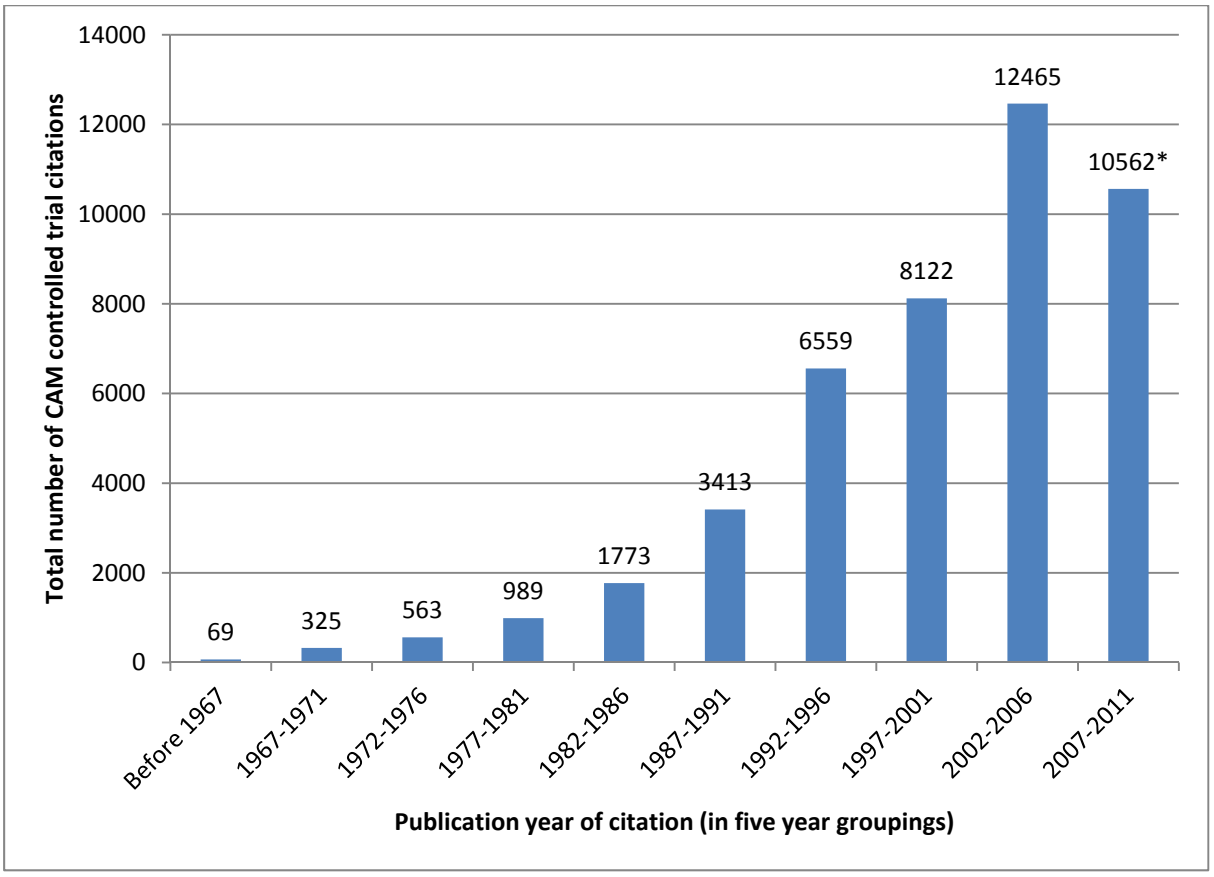
To characterize register citations by medical conditions, we used the 25 categories of medical conditions listed in the browse list on the home page of *The Cochrane Library*.[20] We developed subject searches for each medical condition category by consulting Cochrane reviews grouped under each of these 25 categories to ascertain relevant search terms and concepts, and additional sources to identify any supplementary terms. Searches were run and tested in MEDLINE in the same manner as described above for the CAM intervention searches, although these searches on medical conditions were not peer reviewed. Search strategies were then adapted so that they could be used to search the CAM register and augmented with additional free text synonyms for relevant medical conditions. Both MeSH and free text terms were sought in all reference management database fields.

RESULTS

At the time of our analysis, the CAM Field register of trials included 44,840 citations of CAM trials, which represented approximately 6% of the total number of all trial citations in CENTRAL.[21] Of these 44,840 CAM trial citations, 15,990 (36%) are not included in MEDLINE. Among the random sample of 200 register citations checked for EMBASE status, 63/200 (31.5%; 95% CI 25% to 38%) are not included in EMBASE. Among these 63 citations not included in EMBASE, only 7/63 (11.1 %; 95% CI 3% to 19%) are included in MEDLINE. Among the 69/200 sample citations not included in MEDLINE, only 13/69 (18.8%; 95% CI 10% to 28%) are included in EMBASE. Overall, 56/200 (28%; 95% CI 22% to 34%) of sample citations are not included in either MEDLINE or EMBASE.

There is a substantial increase in the numbers of trial citations published for each 5 year time period, and the majority of the citations are from more recent publication years (see Figure 1). This increase in the numbers of trial citations included in the CAM Field register over publication year time periods corresponds to a similar increase over publication year periods seen overall in CENTRAL.[22]

Figure 1. Number of citations in the CAM Field specialized register by citation publication year



* The drop in the number of trial citations for 2007-2011 may be partially due to incomplete indexing of trials published in this recent time period and/or may be partially due to our having suspended additions of new search results to the CAM Field specialized register in August 2011.

Table 1. Source of citations in the CAM Field specialized register

Source of citations	All citations *n* (% of citations in register)	Non-MEDLINE citations *n* (% of non-MEDLINE citations in register)
Contributing organization		
Beijing University of Chinese Medicine	6,484 (14%)	6,183 (39%)
Ottawa Hospital Research Institute	2,967 (7%)	2,777 (17%)
Japan Society for Oriental Medicine	351 (1%)	293 (2%)
Korea Institute of Oriental Medicine	307 (1%)	304 (2%)
South Asian Cochrane Centre & Network	71 (<1%)	31 (<1%)
South African Cochrane Centre	26 (<1%)	0 (0)
Total citations from contributing organizations	**10,206 (23%)**	**9,588 (60%)**
Total citations from other sources (e.g., searches of CENTRAL)	**34,634 (77%)**	**6,402 (40%)**
Totals	**44,840 (100%)**	**15,990 (100%)**

The most common languages in the register after English are Chinese (*n*=10,376; 23%), German (*n*=963; 2%), Korean (*n*=330; 0.7%), Japanese (*n*=312; 0.7%) and Russian (*n*=227; 0.5%). The representation of citations in languages other than English probably reflects both the number of trials published in that language as well as our methods of sourcing the trial citations for the register. For example, the number of trials in Chinese reflects both the fact that Chinese journals publish a large number of trials of traditional Chinese medicine interventions (e.g., acupuncture, Chinese herbal medicine)[23] and also the fact that we have collaborated with our partner institution in China, since 2008, to search Chinese journals and databases to identify these trials. More recent collaborations have resulted in the identification and inclusion in the register of traditional medicine trial citations from other countries (see Table 1). Searches of trials by several of these collaborating institutions are ongoing.

The 25 conventional journal titles and 25 CAM journal titles with the largest number of citations in the register are listed in Tables 2 and 3, respectively. Nearly all (24/25) of the top conventional medicine journals are MEDLINE-indexed, while only 12/25 of the top 25 CAM journals are MEDLINE-indexed. Together, these 50 journals account for 13,731 trials (30.6% of the total) and 8,445 of these trial citations (61.5%) are MEDLINE-indexed. The clinical focus of the conventional journals was concentrated in nutrition (*n*=8), general and internal medicine (*n*=4), and pediatrics (*n*=3). The clinical focus of the top 25 CAM journals was concentrated in TCM (*n*=20).

The CAM register contains citations from 4,845 journals. Citations are quite concentrated in a few journals. One-third of the citations are found in the top 57 journals and two-thirds come from the top 420 journals. Among the 4,425 journals containing the remaining one-third of citations, 2,749/4,425 (62%) contributed only 1 or 2 citations to the register.

Of the 44,840 trial citations in the register, 93% were classified into one or more of the CAM intervention categories for which searches were conducted. The greatest concentrations were in non-vitamin, non-mineral dietary supplements (e.g., glucosamine, fish oil); Chinese herbal medicine (e.g., *Astragalus membranaceus*, *Schisandra chinensis*); diet-based therapies; vitamin and mineral interventions; and acupuncture (Table 4). The high representation of acupuncture and Chinese herbal medicine trial citations in the register might be explained by both a large number of trials being published in these areas, as well as by the methods that we used to source the trial citations for the database (as described above).

Of the 44,840 citations in the register, 85% were classified into one or more of the 25 categories of medical conditions. The greatest concentrations were in the categories of heart & circulation, anesthesia & pain, mental health, and endocrine & metabolic conditions (Table 5). Categories varied greatly in the proportion of citations included on MEDLINE, and the lowest percentage of MEDLINE-indexed citations was among citations not classified into any medical condition category. This is likely a result of non-MEDLINE citations being less likely to have abstracts or detailed keywords, and thus being less easy to categorize through searches.

Table 2. Twenty-five conventional medicine journals with the most citations in the CAM Field specialized register

Journal name*	Number of citations (% of citations in register)	Language of full text	MED-LINE indexed
The American Journal of Clinical Nutrition	1,027 (2.3)	English	Yes
European Journal of Clinical Nutrition	381 (0.8)	English	Yes
The Journal of Nutrition	362 (0.8)	English	Yes
The British Journal of Nutrition	310 (0.7)	English	Yes
Lancet	247 (0.6)	English	Yes
BMJ	223 (0.5)	English	Yes
Pain	182 (0.4)	English	Yes
Journal of Consulting and Clinical Psychology	178 (0.4)	English	Yes
Behaviour Research and Therapy	170 (0.4)	English	Yes
Journal of Clinical Rehabilitative Tissue Engineering Research [Zhong Guo Zu Zhi Gong Cheng Yan Jiu Yu Lin Chuang Kang Fu]	166 (0.4)	Chinese	No
Archives of Physical Medicine and Rehabilitation	164 (0.4)	English	Yes
Journal of the American College of Nutrition	156 (0.3)	English	Yes
Diabetes Care	144 (0.3)	English	Yes
JAMA : Journal of the American Medical Association	136 (0.3)	English	Yes
Atherosclerosis	132 (0.3)	English	Yes
Journal of Clinical Endocrinology and Metabolism	131 (0.3)	English	Yes
Nutrition (Burbank, Los Angeles County, Calif.)	128 (0.3)	English	Yes
Pediatrics	127 (0.3)	English	Yes
The Journal of Pediatrics	121 (0.3)	English	Yes
Lipids	119 (0.3)	English	Yes
The New England Journal of Medicine	111 (0.2)	English	Yes
Arzneimittel-Forschung	110 (0.2)	English, German†	Yes
Medicine and Science in Sports and Exercise	91 (0.2)	English	Yes
Journal of Pediatric Gastroenterology and Nutrition	90 (0.2)	English	Yes
Journal of the American Dietetic Association	88 (0.2)	English	Yes
Total	5,094 (11.4)	English	Yes

*Totals include named journals and any predecessor journals that were continued by the named journal.
†Articles are published in either German or English, and abstracts are available in both languages.

Table 3. Twenty-five complementary medicine journals with the most citations in the CAM Field specialized register

Journal name*	Number of citations (% of citations in register)	Language of full text	MEDLINE indexed
Chinese Journal of Integrated Traditional and Western Medicine [Zhong Guo Zhong Xi Yi Jie He Za Zhi]	1,809 (4.0)	Chinese	Yes
Chinese Journal of Information on Traditional Chinese Medicine [Zhong Guo Zhong Yi Yao Xin Xi Za Zhi]	1,049 (2.3)	Chinese	No
Chinese Acupuncture & Moxibustion [Zhongguo zhen jiu]	1,021 (2.3)	Chinese	No
Shanghai Journal of Acupuncture and Moxibustion [Shang Hai Zhen Jiu Za Zhi]	557 (1.2)	Chinese	No
Journal of Traditional Chinese Medicine	367 (0.8)	English	Yes
Journal of Alternative and Complementary Medicine	357 (0.8)	English	Yes
Chinese Journal of Integrated Traditional and Western Medicine on Liver Diseases [Zhong Xi Yi Jie He Gan Bing Za Zhi]	356 (0.8)	Chinese	No
Modern Journal of Integrated Traditional Chinese and Western Medicine [Xian Dai Zhong Xi Yi Jie He Za Zhi]	350 (0.8)	Chinese	No
Journal of Manipulative and Physiological Therapeutics	274 (0.6)	English	Yes
Chinese Traditional Patent Medicine [Zhong Cheng Yao]	231 (0.5)	Chinese	No
China Journal of Chinese Materia Medica [Zhong Guo Zhong Yao Za Zhi]	189 (0.4)	Chinese	Yes
Shanghai Journal of Traditional Chinese Medicine [Shang Hai Zhong Yi Yao Za Zhi]	186 (0.4)	Chinese	No
Hebei Journal of Traditional Chinese Medicine [He Bei Zhong Yi]	184 (0.4)	Chinese	No
Chinese Journal of Integrative Medicine	178 (0.4)	English	Yes
The American Journal of Chinese Medicine	169 (0.4)	English	Yes
Shandong Journal of Traditional Chinese Medicine [Shan Dong Zhong Yi Za Zhi]	163 (0.4)	Chinese	No
Complementary Therapies in Medicine	150 (0.3)	English	Yes
World Journal of Acupuncture-Moxibustion	147 (0.3)	English	No
China Journal of Traditional Chinese Medicine and Pharmacy [Zhong Hua Zhong Yi Yao Za Zhi]	146 (0.3)	Chinese	No
Jiangsu Journal of Traditional Chinese Medicine [Jiang Su Zhong Yi Yao]	134 (0.3)	Chinese	No
Acupuncture Research [Zhen Ci Yan Jiu]	133 (0.3)	Chinese	Yes
Journal of Chinese Integrative Medicine [Zhong Xi Yi Jie He Xue Bao]	129 (0.3)	Chinese	Yes
Phytomedicine : International Journal of Phytotherapy and Phytopharmacology	128 (0.3)	English	Yes
Journal of Psychosomatic Research	124 (0.3)	English	Yes
Journal of the Korean Acupuncture & Moxibustion Society [Taehan Chimgu Hakhoe chi]	106 (0.2)	Korean	No
Total	8,637 (19.3)		

*Totals include named journals and any predecessor journals that were continued by the named journal.

Table 4. Number of CAM Field specialized register citations classified by type of CAM interventions

CAM Intervention	Citations *n* (% of citations in register)	MEDLINE indexed citations *n* (% of citations in intervention category that are MEDLINE-indexed)
Non-vitamin, non-mineral dietary supplements (e.g., glucosamine, fish oil)	15,140 (33.8)	12,529 (82.8)
Chinese herbal medicine (e.g., *Astragalus membranaceus*, *Schisandra chinensis*)	12,118 (27.0)	3,575 (29.5)
Diet-based therapies	9,009 (20.1)	9,009 (88.3)
Vitamin and mineral interventions (includes megavitamin therapies and vitamin or mineral therapies for other than medically diagnosed deficiencies or deficiency-related disorders)	7,741 (17.3)	6,468 (83.6)
Acupuncture	6,035 (13.5)	2,632 (43.6)
Relaxation (includes guided imagery and deep breathing)	3,743 (8.3)	3,194 (85.3)
Interventions using veritable energy modalities (unconventional uses of magnets, phototherapy, electrical stimulation, or ultrasonic therapy)	2,977 (6.6)	2,265 (76.1)
Chiropractic or osteopathic manipulation	2,606 (5.8)	2,041 (78.3)
Biofeedback	2,109 (4.7)	1,643 (77.9)
Massage	1,481 (3.3)	987 (66.6)
Traditional medicine not otherwise specified (e.g., Ayurveda, Kampo)	1,409 (3.1)	560 (39.7)
Meditation (includes mindfulness-based therapies)	1,259 (2.8)	1,056 (83.9)
Biologically based interventions not otherwise specified (e.g., balneotherapy, prolotherapy) and excluding interventions using energy fields	1,215 (2.8)	944 (77.7)
Interventions using putative energy fields (distant healing, prayer, qi gong, reiki, spiritual healing, and therapeutic touch)	1,210 (2.7)	801 (66.2)
Sensory art therapies (includes art, dance, drama, music, and play therapy)	1,136 (2.5)	911 (80.2)
Hypnosis	780 (1.7)	600 (76.9)
Homeopathy	755 (1.7)	302 (40.0)
Manipulative and body based therapies not otherwise specified (e.g., Alexander technique, Pilates)	438 (0.9)	212 (48.4)
Yoga	333 (0.7)	242 (72.7)
Tai chi	188 (0.4)	149 (79.3)
Chelation therapy	153 (0.3)	148 (96.7)
Unclassified	3,093 (6.9)	2,124 (68.7)
Totals*	44,840	28,850 (64.3)

*Because individual citations were frequently classified into more than one CAM intervention category, the sum of the numbers of citations classified into each CAM intervention category exceeds the total number of citations in the CAM Field register (i.e., 44,840).

Table 5. Number of CAM Field specialized register citations classified by type of medical condition

Medical conditions	Citations *n* (% of citations in register)	MEDLINE indexed *n* (% of citations in medical condition category that are MEDLINE indexed)
Heart & circulation	8,028 (17.9)	5,585 (69.6)
Anesthesia & pain control	7,656 (17.1)	5,492 (71.7)
Mental health	7,472 (16.7)	5,646 (75.6)
Endocrine & metabolic	6,188 (13.8)	3,797 (61.4)
Gastroenterology	3,378 (7.5)	2,354 (69.7)
Orthopedics & trauma	3,331 (7.4)	2,263 (67.9)
Cancer	2,983 (6.7)	2,205 (73.9)
Lungs & airways	2,545 (5.7)	1,828 (71.8)
Tobacco, drugs, & alcohol dependence	2,286 (5.1)	1,935 (84.6)
Neonatal care	2,252 (5.0)	1,924 (85.4)
Rheumatology	2,199 (4.9)	1,580 (71.9)
Infectious disease	2,080 (4.6)	1,128 (54.2)
Pregnancy & childbirth	2,050 (4.6)	1,642 (80.1)
Kidney disease	1,849 (4.1)	1,207 (65.3)
Neurology	1,697 (3.8)	1,289 (76.0)
Gynecology	1,598 (3.6)	1,142 (71.5)
Skin	1,187 (2.6)	746 (62.8)
Dentistry & oral health	1,123 (2.5)	896 (79.8)
Ear, nose, & throat	1,108 (2.5)	742 (67.0)
Eyes & vision	764 (1.7)	505 (66.1)
Urology	690 (1.5)	593 (85.9)
Wounds	667 (1.5)	505 (75.7)
Developmental, psychosocial, & learning problems	656 (1.5)	567 (86.4)
Blood disorders	596 (1.3)	435 (73.0)
Genetic disorders	214 (0.5)	148 (69.2)
Unclassified	6,635 (14.8)	3,298 (49.7)
Totals*	44,840	28,850 (64.3)

*Because individual citations were frequently classified into more than one medical condition category, the sum of the numbers of citations classified into each medical condition category exceeds the total number of citations in the CAM Field register (i.e., 44,840).

DISCUSSION

The CAM Field specialized register is an important resource for both MEDLINE and non-MEDLINE citations of CAM controlled trials. The prevalence of MEDLINE-indexed trial citations reflects the searches conducted in PubMed (and later in CENTRAL), for MEDLINE-indexed citations retrieved using the CAM on PubMed search strategy. We must therefore acknowledge the strides made in identification of both controlled trial and CAM citations by the US National Library of Medicine since the CAM Field specialized register was last examined in 1998.[5] The large number of non-MEDLINE citations reflects searches of CENTRAL, which includes non-MEDLINE citations, and the efforts of CAM Field collaborators in China, Canada, Japan, and Korea, whose contributions to the specialized register were of predominantly non-MEDLINE citations. Overall, less than two-thirds of register citations are MEDLINE-indexed, and Sampson et al. concluded that with incomplete MEDLINE indexing of a body of literature, a specialized register was of particular utility.[24]

Particular strengths of the register include citations of nutritional and supplement-related interventions, and traditional medicine. While citations of chelation therapy, nutrition and supplement-related interventions, sensory art therapies, relaxation, and meditation are likely to be MEDLINE-indexed (at least 80% of all these citations are indexed in MEDLINE), citations of traditional medicine interventions and homeopathy are less likely to be MEDLINE-indexed (fewer than 50% indexed in MEDLINE) and thus may be more difficult to locate. Therefore, the CAM Field specialized register may be a particularly useful resource for identifying citations of trials to be included in systematic reviews of traditional medicine interventions, particularly TCM. Similarly, the CAM Field specialized register may be a useful source of trials for systematic reviews covering CAM interventions for medical conditions in which a lower proportion of citations are MEDLINE-indexed (e.g., endocrine & metabolic conditions or infectious disease).

Among the 4,845 journals containing citations in the CAM Field register, 9% contain two-thirds of the register citations, and 57% contain only one or two register citations each. One unanswered question is whether this distribution of journals in the register represents the *true* distribution of CAM trials across journals or whether instead it is an artifact resulting from the way that the register was developed. There is no way to definitively answer this question because there exists no 'gold standard' complete database of CAM trials against which the journal distribution in the CAM Field register can be compared. However, we believe that the distribution of journals in the CAM Field register is largely an artifact of the way the register was developed. This is because, in identifying citations for register inclusion, contributors often searched bibliographic databases to identify trial citations on specific topics for their systematic reviews, rather than comprehensively searching entire journals for all CAM trials. As a result, some journals for which only one or two CAM trial citations were identified for register inclusion may have many additional trials that have not yet been identified. Continued efforts to identify trials will likely change the distribution of citations across journals. Such efforts may also change the characteristics of the register in other ways (e.g., the number and proportion of trials covering particular CAM interventions or published in particular languages) that are not possible to quantify in advance. While the ultimate aim is for the CAM Field register to be a comprehensive source of CAM controlled trials, the register cannot currently be considered to be comprehensive. Therefore, systematic reviewers of CAM interventions should search multiple electronic and other sources for relevant CAM trials, in addition to searching the CAM Field register.

The strength of the register in terms of its coverage of difficult to locate trials may, however, be associated with potential weaknesses in terms of the quality of these trials. The largest subset of non-MEDLINE citations in the register (51%) is trials published in Chinese. These Chinese-language trials were included in the register if the trial publication stated that a random or quasi-random

procedure was used to assign participants to treatment groups. However, a recent telephone survey of authors of 'claimed' randomized trials conducted in China discovered that only 7% could be confirmed to use a random method to assign participants to treatment groups.[25] Inclusion in systematic reviews of such Chinese trials claiming to be randomized, but not confirmed as such by systematic reviewers, may inflate these reviews' meta-analytic effect estimates.[26] In addition, a 1998 review of the outcomes of non-English language trials by Vickers et al found that acupuncture trials conducted in China reported positive results 100% of the time, and Chinese trials of other interventions reported positive results 99% of the time, strongly suggesting the preferential publication in China of trials with positive results.[1,27] Although Chinese language trials reflect the majority of non-English language trials included in the register, the issue of a publication or reporting bias favoring positive results may also be relevant to other non-English trials included in the register. For example, the Vickers et al 1998 review found that not only acupuncture trials from China, but also acupuncture trials from Japan, Hong Kong, and Taiwan were uniformly positive. Also, a recent preliminary investigation into the results of the Japanese Kampo trials included in the CAM Field register found that only a small number of Kampo trials have negative results [Kiichiro Tsutani personal communication, 2011 March 14], and an informal assessment of CAM Korean trials indicated that negative results were rare [Byeung-Cheul Shin personal communication, 2012 July 13]. The positive results of non-English language CAM trials likely explain why including such trials in CAM-related systematic reviews tends to inflate meta-analysis effect estimates, according to a 2005 empirical study.[28]

It is possible that the methodological quality of the more recent Chinese-language trials may, however, be better than that of the earlier trials because, for example, the CONSORT statement has recently been more widely disseminated in China,[29] including in Chinese journals of TCM.[30] In addition, while the Vickers et al 1998 review[27] found that Chinese acupuncture trials published up to 1998 were uniformly positive, it is not known whether or not more recent Chinese trials are also likely to be uniformly positive. An important topic area for future research is to determine whether there is a publication bias favoring positive results in more recent trials from China. However, while conducting research studies to assess for the likelihood of publication bias in Chinese trials may be informative in determining the scope of the problem, the only way to avoid publication bias in Chinese trials is to ensure that all initiated Chinese language trials are known about through the registration of Chinese trials at inception, which is currently being implemented.[31] Universal trial registration, in conjunction with reporting of trial registration numbers in publications, might also serve as a tool in addressing duplicate publication,[32] which some studies have observed to be prevalent among Chinese, Japanese, and Korean trials.[33-35] In the interim, for systematic reviews including a large number of Chinese trials, a possible approach for assessing the impact of a potential publication bias related to the Chinese trials may be to mark the Chinese trials in funnel plots in systematic reviews.

Despite these concerns over the validity of Chinese trials in general, it seems inappropriate to exclude trials from systematic reviews on the basis of language or country of publication alone. A more measured approach may be to search for Chinese trials and to telephone interview the authors of potentially eligible trials to try and assess whether the trials were truly randomized before including them in the review. If concerns about the validity of the trials remain, even after the telephone interviews with the authors, a possible approach is to include in the review those trials (Chinese or Western) for which there remains uncertainty about whether true randomization was used, but to be more restrictive when presenting the key findings, such as the abstract conclusions and the summary of findings table.[36] Another approach may be to analyze the potential influence of risk of bias measures (e.g., adequacy of randomization) on effect estimates using subgroup analyses or sensitivity analyses. Either way, such assessment and analysis approaches should probably be

based on risk of bias measures rather than on the language of country of origin of the trials. This is because generalizing about *individual* Chinese language trials, for example, based on meta-research of the characteristics of Chinese trials *overall* would be an erroneous oversimplification. Instead, each trial included in a systematic review needs to be individually evaluated on its risks of bias, assuming either that the trial publication is sufficiently informative, or that the trial author can be contacted for further information. Such risk of bias assessments can then be incorporated into the review's analysis.

In addition to providing a source of trials for inclusion in systematic reviews, the register may also be used for investigations into the optimal use of CAM research resources and the prioritization of future CAM reviews. This analysis of the types of CAM interventions and health conditions covered in the register is a first step in conducting such investigations. That is, the number of citations related to various CAM interventions (e.g., diet-based therapies) and the number of citations related to various medical conditions (e.g., endocrine & metabolic disorders) may be used for research into whether there is a correlation between those CAM interventions most frequently investigated in trials and those most commonly used, and whether the most serious or prevalent health conditions are proportionately represented with the highest number of CAM trials. If trials of commonly used interventions and/or trials for serious or prevalent health conditions are lacking, this may indicate that CAM research resources should be directed to these areas. In addition, the trial database may also be useful for prioritizing future Cochrane reviews by identifying CAM intervention/health condition pairings for which there are available trials in the register but no existing Cochrane review. The fact that the same classification categories were used for CAM intervention types and health conditions, for both the trial citations in the specialized register and for a separate database of CAM-related Cochrane reviews,[10] should facilitate such identification. However identifying potential future systematic reviews to prepare will require additional narrowing down of some of our CAM intervention type categories (e.g., "Chinese herbal medicine") and health condition categories (e.g., "mental health") in order to identify more specific intervention/condition pairings (e.g., the Chinese herbal medicine formula Free and Easy Wanderer for depression) for systematic reviews. Future plans for the CAM Field specialized register include augmenting the size and scope of the register through ongoing searches and international partnerships, and developing methods to characterize groups of trials according to intervention/condition characteristics and mapping these groups to gaps in Cochrane systematic reviews.

CONCLUSIONS

The number of citations included in the CAM Field specialized register increased nearly tenfold between 2006 and 2011 as a result of a program of extensive searching and partnerships with international collaborators. Many CAM Field register citations are not MEDLINE-indexed and many of these non-MEDLINE-indexed citations are published in languages other than English. While the register provides access to thousands of difficult to locate citations of trials, many of these trials are likely to be of low quality and may overestimate treatment effects. When these trials are considered for inclusion in systematic reviews, it is extremely important that their risk of bias is adequately assessed.

REFERENCES

1. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002;25(1):38-64.
2. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7(1):1-76.
3. Hofmans EA. Acupuncture and Medline. *Lancet* 1990;336(8706):57.
4. Kleijnen J, Knipschild P. The comprehensiveness of Medline and Embase computer searches. Searches for controlled trials of homoeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication. *Pharm Weekbl Sci* 1992;14(5):316-20.
5. Vickers AJ. Bibliometric analysis of randomized trials in complementary medicine. *Complementary Therapies in Medicine* 1998;6:185-9.
6. Lefebvre C, Manheimer E, Glanville J (editors). Chapter 6: Searching for studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (accessed 2013 July 12).
7. Committee on the Use of Complementary and Alternative Medicine by the American Public. Chapter 5: State of emerging evidence on CAM. In: Board of Health Promotion and Disease Prevention, Institute of Medicine of the National Academies. *Complementary and Alternative Medicine in the United States*. Washington DC: The National Academies Press; 2005. Available from http://books.nap.edu/catalog/11182.html (accessed 2013 July 12).
8. US National Center for Complementary and Alternative Medicine. About CAM on PubMed. Bethesda, MD: US Department of Health and Human Services. National Institutes of Health. Available from http://nccam.nih.gov/research/camonpubmed (accessed 2013 July 12).
9. US National Library of Medicine. Search strategy used to create the Complementary Medicine subset on PubMed. Bethesda, MD: US Department of Health and Human Services. National Institutes of Health. Available from http://www.nlm.nih.gov/bsd/pubmed_subsets/comp_med_strategy.html (accessed 2013 July 12).
10. Wieland LS*, Manheimer E*, Berman BM. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane collaboration. *Altern Ther Health Med* 2011;17(2):50-9. *Both authors contributed equally.
11. World Health Organization. Traditional medicine, Fact Sheet No. 134. Geneva: World Health Organization; 2008. Available from http://www.who.int/mediacentre/factsheets/fs134/en/index.html (accessed 2013 July 12).
12. Cogo E, Sampson M, Ajiferuke I, Manheimer E, Campbell K, Daniel R, Moher D. Searching for Controlled Trials of Complementary and Alternative Medicine: a Comparison of 15 Databases. *Evid Based Complement Alternat Med* 2011 June 23;858246.
13. Kim KH, Kong JC, Choi JY, Choi TY, Shin BC, McDonald S, Lee MS. Impact of including Korean randomized controlled trials in Cochrane reviews of acupuncture. *PLoS One* 2012;7(10):e47619.
14. Okabe T, Tsutani K. Evidence Reports of Kampo Treatment 2010:345 Randomized Controlled Trials (EKAT 2010).Task Force for Evidence Reports / Clinical Practice Guidelines (ER/CPG-TF), Special Committee for Evidence-based Medicine (EBM),The Japan Society for Oriental Medicine (JSOM), 2011. Available from http://www.jsom.or.jp/medical/ebm/ere/pdf.html (accessed 2013 July 12).
15. Barnabas J, Yamuna BGC, Parthasarathy V, Venkatesh S, Tharyan P. Access to evidence from countries in South Asia: the South Asian Database of Controlled Clinical Trials and the South Asian Cochrane Network and Centre's Digital Library. In: 17th International Cochrane Colloquium; 2009 Oct 11-14; Singapore. Abstract no. P06-8. Available at http://www.imbi.uni-freiburg.de/OJS/cca/index.php?journal=cca&page=article&op=view&path%5B%5D=8010 (accessed 2013 January 14).
16. Pienaar E. African Trials Register: 5 years later, where are we now? In: 13th International Cochrane Colloquium; 2005 Oct 22-26; Melbourne. Abstract no. P021. Available at http://www.imbi.uni-freiburg.de/OJS/cca/index.php?journal=cca&page=article&op=view&path%5B%5D=1232 (accessed 2013 January 14).
17. Siegfried N, Busgeeth K, Clarke M, Pienaar E, Volmink L. African randomized controlled trials - too hard to find? In: 14th International Cochrane Colloquium; 2006 October 23-26; Dublin. Abstract no. P017. Available at http://www.imbi.uni-freiburg.de/OJS/cca/index.php?journal=cca&page=article&op=view&path%5B%5D=1972 (accessed 2013 January 14).
18. Barnes PM, Bloom B, Nahin RL. Complementary and alternative medicine use among adults and children: United States, 2007. *Natl Health Stat Report* 2008(12):1-23.
19. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J Clin Epidemiol* 2009;62(9):944-52.
20. The Cochrane Library: Independent high quality evidence for healthcare decision making. John Wiley & Sons, Ltd. Available from http://www.thecochranelibrary.com/view/0/index.html (accessed 2013 July 12).

21. Cochrane Central Register of Controlled Trials (CENTRAL) 2012, Issue 1, part of *The Cochrane Library*. www.thecochranelibrary.com.
22. Clarke M, Clarke T. The Cochrane Central Register of Controlled Trials and the contribution from the specialized registers of Cochrane Review Groups. 2008. A project funded in part by a grant from The Cochrane Collaboration's Disretionary Fund.
23. Tang JL, Zhan SY, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;319(7203):160-1.
24. Sampson M, Daniel R, Cogo E, Dingwall O. Sources of evidence to support systematic reviews in librarianship. *J Med Libr Assoc* 2008;96(1):66-9.
25. Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials* 2009;10:46.
26. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157(6):429-38.
27. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;19(2):159-66.
28. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. *J Clin Epidemiol* 2005;58(8):769-76.
29. MacPherson H, Altman DG. Improving the quality of reporting acupuncture interventions: describing the collaboration between STRICTA, CONSORT and the Chinese Cochrane Centre. *J Evid Based Med* 2009;2(1):57-60.
30. Fei YT, Liu JP. [Improving the quality of reporting Chinese herbal medicine trials: an elaborated checklist]. *Zhong Xi Yi Jie He Xue Bao* 2008;6(3):233-8.
31. Wu T, Li Y, Liu G, Li J, Wang L, Du L. Chinese clinical trial registry: mission, responsibility and operation. *J Evid Based Med* 2011 Aug 2. doi: 10.1111/j.756-5391.2011.01137.x. [Epub ahead of print]
32. Antes G, Dickersin K. Trial registration to prevent duplicate publication. *JAMA* 2004;291(20):2432.
33. Errami M, Garner H. A tale of two citations. *Nature* 2008;451(7177):397-9.
34. Kim SY, Hahm CK, Bae CW, Cho HM. Duplicate Publications in Korean medical journals indexed in KoreaMed. *J Korean Med Sci* 2008;23(1):131-3.
35. Kitagawa M, Tsutani K. Duplicate publication cases in the field of Kampo (Japanese herbal medicine) in Japan. *Zhong Xi Yi Jie He Xue Bao* 2011;9(10):1055-60.
36. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (accessed 2013 July 12).

Chapter 4. Evidence from the Cochrane Collaboration for traditional Chinese medicine therapies

Manheimer E, Wieland S, Kimbrough E, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for traditional Chinese medicine therapies. J Altern Complement Med 2009 Sep;15(9):1001-14.

ABSTRACT

BACKGROUND: The Cochrane Collaboration, an international not-for-profit organization that prepares and maintains systematic reviews of randomized trials of healthcare therapies, has produced reviews summarizing much of the evidence on traditional Chinese medicine. Our objective was to review the evidence base for traditional Chinese Medicine (TCM) according to Cochrane systematic reviews.

METHODS: In order to detect reviews focusing on TCM, we searched the titles and abstracts of all reviews in Issue 4, 2008 of the Cochrane Database of Systematic Reviews. For each review, we extracted data on the number of trials included and the total number of participants. We provided an indication of the strength of the review findings by assessing the reviewers' abstract conclusions statement. We supplemented our assessment of the abstract conclusions statements with a listing of the comparisons and outcomes showing statistically significant meta-analyses results.

RESULTS: We identified 70 Cochrane systematic reviews of TCM, primarily acupuncture (n=26) and Chinese herbal medicine (n=42), and 1 each of moxibustion and tai chi. 19/26 acupuncture reviews and 22/42 herbal medicine reviews concluded that there was not enough good quality trial evidence to make any conclusion about the efficacy of the evaluated treatment, while the remaining 7 acupuncture and 20 herbal medicine reviews and each of the moxibustion and tai chi reviews indicated a suggestion of benefit, which was qualified by a caveat about the poor quality and quantity of studies. Most reviews included many distinct interventions, controls, outcomes, and populations, and a large number of different comparisons were made, each with a distinct forest plot.

CONCLUSIONS: Most Cochrane systematic reviews of TCM are inconclusive, due specifically to the poor methodology and heterogeneity of the studies reviewed. Some systematic reviews provide preliminary evidence of Chinese medicine's benefits to certain patient populations, underscoring the importance and appropriateness of further research. These preliminary findings should be considered tentative and need to be confirmed with rigorous randomized controlled trials.

The online version of the thesis Table of Contents includes links to the following Cochrane Colloquia abstracts, which were preliminary work for this chapter:

- Manheimer E, Kimbrough E, Wieland S, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for traditional Chinese medicine therapies. In: 17th International Cochrane Colloquium; 2009 Oct 11-14; Singapore. p. 58-59. Abstract no. P85.
- Manheimer E, Berman B, Dubnick H, Beckner W. Cochrane reviews of complementary and alternative therapies: evaluating the strength of the evidence. In: 12th International Cochrane Colloquium; 2004 Oct 2-6; Ottawa. p. 159-160. Abstract no. P-094.

BACKGROUND

The Use of Traditional Chinese Medicine

Traditional Chinese medicine (TCM) approaches include acupuncture, herbal medicine, moxibustion and Tai Chi/Qigong. Acupuncture and herbal medicine are two of the most commonly used complementary medicine therapies.[1,2] In a nationally-representative US survey conducted in 2002, almost 20% of adults had used herbal therapies in the past year,[3] and about 1% had used acupuncture.[4] Among Asian-Americans, the prevalence was much higher, with 75-100% reporting using TCM, primarily herbal medicine.[2] The allure of these therapies is not driven by dissatisfaction with conventional medicine, according to recent evidence,[1,5,6] but rather by the belief that combining Chinese and conventional medicine provides more optimal healing than conventional medicine alone.[1,4,7,8] Chinese medicine users also enjoy a sense of participation in their own healing,[9] and feel a congruence between Chinese medicine and their personal values and philosophical orientation.[5,9,10] The use of Chinese medicine therapies is more prevalent among women and those of higher socio-economic status,[3,7-9,11] with poorer health and higher education levels being additional predictors of its use.[4,8-11] Health concerns driving the use of TCM therapies are primarily those of chronic pain, musculoskeletal problems and mood disturbances, including back and neck pain, joint pain and stiffness, headaches, anxiety and depression.[1,2,4,8,9] TCM users report improvements in both specific symptoms and overall quality of life, according to several surveys.[9]

The Evidence Base for Traditional Chinese Medicine: The Cochrane Collaboration and Systematic Reviews of TCM

Increasing use of TCM approaches in the United States contributed to the establishment in 1998 of the National Center for Complementary and Alternative Medicine (NCCAM) within the US National Institutes of Health. This led to a dramatic expansion in the number and quality of randomized controlled trials (RCTs) funded to study TCM, many of which have focused on acupuncture and traditional Chinese herbal therapies. Concurrent with these trials, systematic reviews have been conducted in order to assess the current state of the research evidence on TCM therapies, and to provide direction for future research.

The independent and not-for-profit Cochrane Collaboration is the world's leading producer of up-to-date systematic reviews of healthcare therapies.[12] The Collaboration currently involves more than 15,000 contributors, most of whom are volunteers, from over 100 countries.[13] The structure of the Collaboration is based upon Cochrane Review Groups, which perform systematic reviews on specific healthcare problems (e.g., schizophrenia, breast cancer), Cochrane Fields, which focus on populations, interventions, or other health care areas that are related to multiple Cochrane Review Groups (e.g., child health, complementary and alternative medicine), and Cochrane Centers, which provide support and assistance to Cochrane Review Groups and Fields in specific countries or geographical areas (e.g, the Australasian Cochrane Centre). The principal product of the Cochrane Collaboration, the *Cochrane Database of Systematic Reviews (CDSR)*, included 3625 complete systematic reviews and 1921 review protocols, as of Issue 4, 2008. Because the Cochrane Collaboration relies on grants and donations and does not accept conflicted funding,[14] Cochrane reviews are largely free from competing financial interests that have the potential to distort and exaggerate findings. The Cochrane Collaboration has overseen a remarkable growth in the evidence base for Chinese medicine therapies. Much of this work has been conducted by the Chinese Cochrane Center and the Cochrane Complementary and Alternative Medicine (CAM) Field. The Cochrane CAM Field is based at the University of Maryland School of Medicine, Center for Integrative Medicine. One of the functions of this Field is to develop databases of all CAM-related Cochrane reviews, classify these reviews according to the CAM treatment modality

investigated (e.g. acupuncture, Chinese herbal medicine), and disseminate information about the reviews to CAM researchers, providers, and consumers worldwide.

METHODS

To determine the current extent of the Cochrane coverage of TCM therapies, we identified each Cochrane review whose primary focus was Chinese herbal medicine, acupuncture, moxibustion, Tai Chi, Qigong, or Tui Na (Chinese massage). In order to detect reviews focusing on TCM, we searched the titles and abstracts of all reviews in Issue 4, 2008 of the CDSR using the text word search strategy in Box 1. We read the title and abstract of each retrieved review in order to confirm that the review was focused on TCM, to identify reviews that have been withdrawn from publication (e.g., because the review is no longer considered current), and to classify the TCM topic of the review. We did not include reviews that stated their focus was on all CAM therapies, on general herbal interventions, or on any other therapies that were not listed above or otherwise identified as specifically Chinese. For example, in the Chinese herbal medicine category, we included only those reviews of herbal medicine used traditionally as TCM, and excluded reviews of individual herbs primarily used in the West. Although ginkgo biloba is sometimes used in the West, it is also commonly used in Chinese medicine, either alone or in combination with other herbs, and therefore we included ginkgo reviews.

Box 1. Text word search strategy applied to title and abstract of systematic reviews in Cochrane Database of Systematic Reviews (CDSR) to identify reviews of TCM

(acupuncture) OR (moxibustion) OR (herb* AND medic*) OR (ginkgo) OR ("tai chi" OR "tai ji") OR (qigong OR "ch'i kung") OR ("tui na") OR (Chinese)

For each review, we extracted data on the number of trials included and the total number of participants. As an indicator of study quality, we also extracted data on the number of trials using adequate versus inadequate/unclear allocation concealment. We used allocation concealment as an indicator of the quality of the included trials for two reasons: first, unconcealed allocation is the study defect shown to be most strongly associated with exaggerated treatment effects,[15] and second, allocation concealment is the only quality measure which is reported in a standardized way for all trials in a Cochrane review (i.e., as adequate, unclear, or inadequate).

In Cochrane reviews that include many distinct interventions, controls, outcomes, and populations, a large number of different comparisons will be made, each with a distinct forest plot. (Forest plots are used to graphically present outcomes of the comparison between a given intervention and a given control for a specific outcome.) To indicate the number of comparisons in a review, and hence its complexity, as well as to show the proportion of comparisons with positive findings, we also extracted the following data for each review: number of forest plots; number of forest plots with a meta-analysis (i.e., statistical pooling of the trial data in the forest plot); and the number of meta-analyses showing a statistically significant benefit of the intervention relative to a control. We included the number of meta-analyses, as conduct of a meta-analysis suggests sufficient homogeneity between the studies under review to allow their results to be quantitatively synthesized.

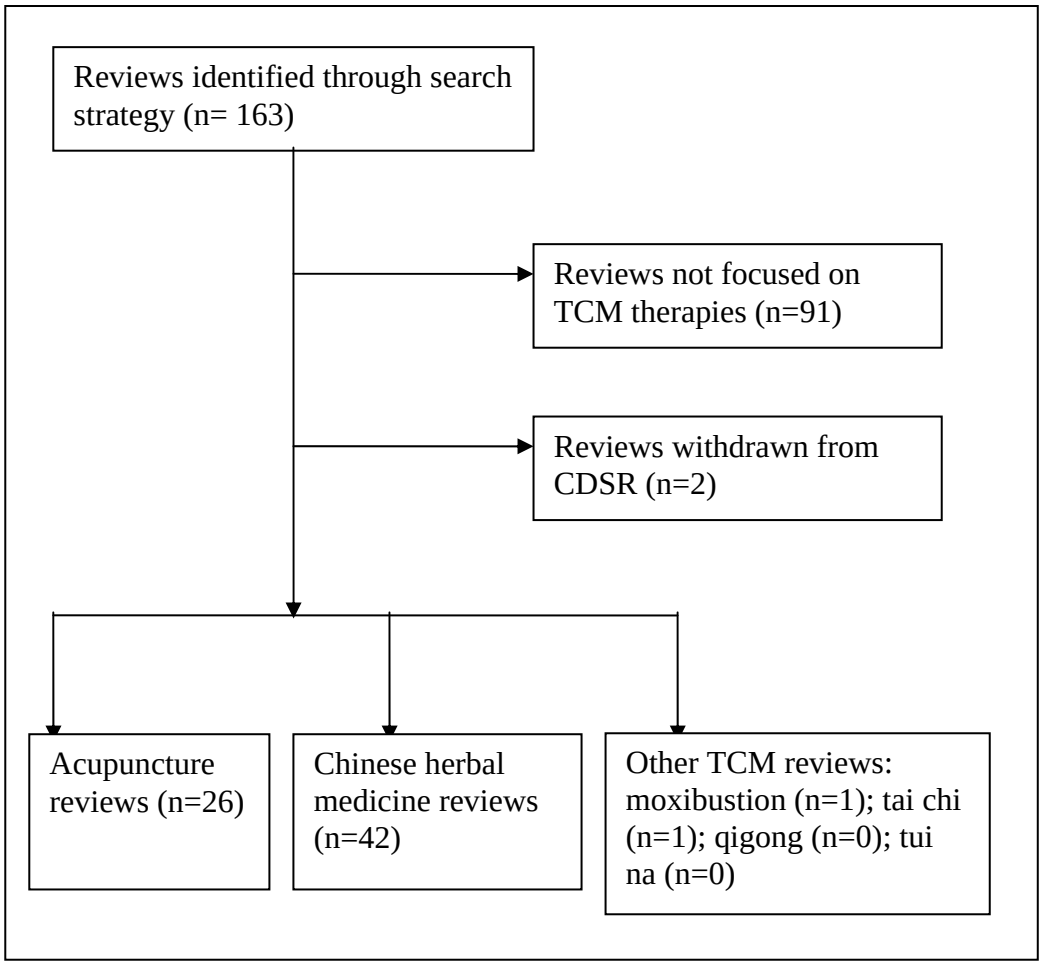
In addition to identifying, listing, and providing descriptive and analytic data about the individual TCM reviews, we also wanted to provide an indication of the nature of the review findings by an assessment of the reviewers' abstract conclusions statement. In Cochrane reviews, the authors' conclusions are statements crafted by the review authors to represent the best overall

assessment of a treatment's effect. These statements are declared only after consideration not only of the findings, but of the nuanced factors contributing to them, including risk of bias, as well as heterogeneity of design, setting, dosage, duration, and outcome measurement among the studies analyzed within the reviews. The Cochrane reviewers attempt to find optimal wording to summarize this complexity for the busy reader who is seeking a brief synopsis of whether or not the treatment 'works'. Editors of Cochrane Review Groups recognize the paramount importance of these statements, and methodically examine them during the peer review process to insure that they accurately reflect the bottom line message. Given the thoughtfulness imbued in the abstract conclusion statement, we felt this would provide an informative reflection of the review. For each review, we assigned the abstract conclusions statement into one of the following two categories: A) the statement indicated a suggestion of benefit, which was qualified by a caveat about the poor quality and quantity of included studies, or B) the statement indicated that the currently available data do not allow any conclusions to be drawn. As an example of an 'A' category review, the 'Herbal medicines for treatment of irritable bowel syndrome' review concluded "Some herbal medicines may improve the symptoms of irritable bowel syndrome. However, positive findings from less rigorous trials should be interpreted with caution due to inadequate methodology, small sample sizes, and lack of confirming data. Some herbal medicines deserve further examination in high-quality trials." As an example of a 'B' category review, the 'Acupuncture for stroke rehabilitation' review concluded "Currently there is no clear evidence on the effects of acupuncture on subacute or chronic stroke. Large methodologically-sound trials are required." Both 'A' and 'B' category reviews generally noted that additional methodologically sound trials were needed before definitive conclusions could be reached. Because the distinction between the 'A' and 'B' reviews was not always obvious, this assignment was carried out by two reviewers making independent assessments. In the case of disagreement between the two, a consensus was reached by discussion; this occurred in 5/70 (7%) of cases. For the reviews classified as category A, we listed the outcomes within each review that showed statistically significant meta-analyses. Also for reviews classified as A, all review data was dually extracted by two authors with consensus reached by discussion in cases of disagreement; for reviews classified as B, one author extracted all data.

RESULTS

Figure 1 shows the results of our search of the CDSR. We identified 70 Cochrane systematic reviews of TCM, primarily acupuncture and Chinese herbal medicine. Table 1 lists the 26 reviews of acupuncture. Table 2 lists the 42 reviews of Chinese herbal medicine, and Table 3 lists the 2 reviews of other TCM therapies (i.e., 1 each of moxibustion and Tai Chi; 0 reviews of Qigong or Tui Na were identified).

Figure 1. Results of searching Cochrane Database of Systematic Reviews (CDSR) for traditional Chinese medicine (TCM) reviews



Acupuncture

We identified 26 Cochrane reviews dedicated to assessing the efficacy of acupuncture (see Table 1). Five reviews focused on treatment of pain (low back pain, shoulder pain, lateral elbow pain, idiopathic headache, and neck disorders), and five on mental or addiction disorders (depression, schizophrenia, vascular dementia, cocaine addiction and smoking cessation). Three reviews focused on gynecological or pregnancy-associated conditions (primary dysmenorrhoea, assisted conception and induction of labor), three reviews focused on stroke (acute stroke, dysphagia in acute stroke, and stroke rehabilitation), and two reviews addressed nausea or vomiting (chemotherapy-induced or postoperative). There was one acupuncture review for each of the following conditions: arthritis, asthma, Bell's palsy, epilepsy, glaucoma, irritable bowel syndrome, and restless legs syndrome.

2/26 (8%) acupuncture reviews were empty reviews; that is, the authors could not find any randomized controlled trials on the relevant topic. Of the remaining reviews, 7 (29%) suggested that there might be some benefit, which was qualified by a caveat about the poor quality and quantity of studies (category A). For 2 of the 7 reviews in category A (29%), the review authors did not conduct any meta-analyses and their conclusion statements were based on a review of the individual trials.[16,17] Meta-analyses were conducted in the remaining 5 (71%). The conditions for which benefit was found for acupuncture were vomiting, nausea, back and neck pain, headache, and assisted conception.

17/24 (71%) of the non-empty reviews concluded that there was not enough good quality trial evidence to make any conclusion (as yet) about the efficacy of the evaluated treatment (category B). In 7/17 (41%) of the reviews in which no benefit from treatment could be conclusively established, significant results were found for at least one meta-analytic outcomes.[18-24] However, the review authors concluded that either the quality of the trials was not sufficient to support a judgment of benefit,[18-20,23,24] that the endpoints were clinically unimportant,[21] or that the heterogeneity[19,22,23] or small sample sizes[19,20,25] precluded drawing conclusions of a benefit. Table 1 shows the number of significant meta-analyses and, for the A category reviews, the outcomes assessed in these meta-analyses.

Table 1. Characteristics of Cochrane reviews of acupuncture

Review title, author and date	Author conclu-sions*	Trials N	Partici-pants N	Trials with allocation conceal-ment N (%)	Forest Plots N	Meta-analyses N	Signifi-cant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Acupuncture and assisted conception. Cheong et al. 2008[26]	A	13	2491	10(77)	24	23	7 (30)	Increased rates of clinical pregnancy, ongoing pregnancy, and live birth
Acupuncture and electroacupuncture for the treatment of rheumatoid arthritis. Casimiro et al., 2005[27]	B	2	84	2 (100)	7	0	N/A	N/A
Acupuncture and dry-needling for low back pain. Furlan et al., 2005[28]	A	35	2861	14 (40)	93	19	15 (79)	Reduced pain and improved functional status
Acupuncture and related interventions for smoking cessation. White et al., 2006[22]	B	24	4949	2 (8)	19	18	5 (28)	N/A
Acupuncture for acute stroke. Zhang et al., 2005[25]	B	14	1208	5 (36)	19	18	5 (28)	N/A
Acupuncture for Bell's palsy. He et al., 2007[29]	B	6	537	0 (0)	0	0	N/A	N/A
Acupuncture for chronic asthma. McCarney et al., 2004[30]	B	12	350	2 (17)	23	2	0 (0)	N/A
Acupuncture for depression. Smith and Hay, 2005[31]	B	7	517	1 (14)	12	4	0 (0)	N/A
Acupuncture for dysphagia in acute stroke. Xie et al., 2008[32]	B	1	66	1 (100)	3	0	N/A	N/A
Acupuncture for epilepsy. Cheuk and Wong, 2008[18]	B	11	914	0 (0)	38	15	5 (33)	N/A
Acupuncture for glaucoma. Law and Li, 2007[33]	B	0	0	N/A	N/A	N/A	N/A	N/A

Review title, author and date	Author conclu-sions*	Trials N	Partici-pants N	Trials with allocation conceal-ment N (%)	Forest Plots N	Meta-analyses N	Signifi-cant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Acupuncture for idiopathic headache. Melchart et al., 2001[16]	A	26	1151	4 (15)	13	0	N/A	N/A
Acupuncture for induction of labor. Smith and Crowther, 2004[21]	B	3	212	3 (100)	8	1	1 (100)	N/A
Acupuncture for insomnia. Cheuk et al., 2007[19]	B	7	590	0 (0)	46	3	2 (67)	N/A
Acupuncture for lateral elbow pain. Green et al., 2002[34]	B	4	239	0 (0)	9	0	N/A	N/A
Acupuncture for neck disorders. Trinh et al., 2006[17]	A	10	661	2 (20)	21	0	0	N/A
Acupuncture for restless legs syndrome. Cui et al., 2008[35]	B	2	170	0 (0)	6	0	N/A	N/A
Acupuncture for schizophrenia. Rathbone and Xia, 2005[20]	B	5	270	0 (0)	11	2	1 (50)	N/A
Acupuncture for shoulder pain. Green et al., 2005[36]	B	9	525	3 (33)	30	1	0 (0)	N/A
Acupuncture for stroke rehabilitation. Wu et al., 2006[23]	B	5	368	0 (0)	1	1	1 (100)	N/A
Acupuncture for treatment of irritable bowel syndrome. Lim et al., 2006[37]	B	6	401	1 (17)	16	1	0 (0)	N/A
Acupuncture for vascular dementia. Peng et al., 2007[38]	B	0	0	N/A	N/A	N/A	N/A	N/A
Acupuncture-point stimulation for chemotherapy-induced nausea or vomiting. Ezzo et al., 2006[39]	A	11	1247	4 (36)	17	13	4 (24)	Reduced risks of vomiting and nausea during first 24 hours
Auricular acupuncture for cocaine dependence. Gates et al., 2006[40]	B	7	1433	1 (14)	10	5	0 (0)	N/A
Stimulation of the wrist acupuncture point P6 for preventing postoperative nausea and vomiting. Lee et al., 2004[41]	A	26	3347	0 (0)	34	24	11 (46)	Reduced risks of nausea, vomiting, and rescue antiemetics

Review title, author and date	Author conclu-sions*	Trials N	Partici-pants N	Trials with allocation conceal-ment N (%)	Forest Plots N	Meta-analyses N	Signifi-cant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Transcutaneous electrical nerve stimulation and acupuncture for primary dysmenorrhoea. Proctor et al., 2002[42]	A	9	220	1 (11)	25	3	1 (33)	Improved overall pain relief

*A = abstract conclusions statement indicated a suggestion of benefit, which was qualified by a caveat about the poor quality and quantity of included studies, and B = abstract conclusions statement indicated that the currently available data do not allow any conclusions to be drawn

†Outcomes are reported only for reviews which 1) include a significant meta-analysis, and 2) for which the authors' abstract conclusions statement indicated a suggestion of benefit

Chinese herbal medicine

We identified 42 Cochrane reviews dedicated specifically to assessing the efficacy of Chinese herbal medicine (see Table 2). Eight reviews focused on treatment of stroke (acute ischemic stroke, acute cerebral infarction), six on treatment of heart problems (angina pectoris, acute myocardial infarction, viral myocarditis, heart failure), five on treatment of mental disorders (schizophrenia, cognitive impairment and dementia, Alzheimer's disease), four on treatment of respiratory problems (severe acute respiratory syndrome, acute bronchitis, common cold), and three each upon treatment of hepatitis (asymptomatic hepatitis B carriers, chronic hepatitis B, hepatitis C virus infection) and gynecological or pregnancy problems (pre-eclampsia, ectopic pregnancy, dysmenorrhoea). The remaining thirteen reviews focused on a range of problems, from lung cancer to tinnitus (see titles of reviews in Table 2). Twenty reviews were focused on single herbs or herbal preparations, while the remainder concerned multiple Chinese herbs or multiple formulations of Chinese herbs.

Of the 42 Cochrane reviews on Chinese herbal medicine, 5/42 (14%) were empty. Of the remaining reviews, 20/37 (54%) of the authors' abstract conclusions supported the possible efficacy of Chinese herbal formulas for treating specific health conditions (category A), while 17/37 (46%) failed to find enough good quality trial evidence to say whether or not the treatment was possibly effective (category B). Of the reviews in category A, 5/20 (25%) did not contain any meta-analyses and the conclusions were based on a review of individual trials. The conditions for which possible benefit was found for Chinese herbal medicine were atopic eczema, primary dysmenorrhoea, schizophrenia, nephritic syndrome, angina, type II diabetes mellitus, severe acute respiratory syndrome, acute pancreatitis, hepatitis B, common cold, side effects of chemotherapy in breast cancer, irritable bowel syndrome, viral myocarditis, Alzheimer's disease, ischaemic stroke, and heart failure.

Of the non-empty reviews for which no benefit from treatment could be conclusively established (i.e., category B reviews), 7/17 (41%) showed statistically significant results from a meta-analysis. In these cases, the review authors concluded that either the quality of the trials was not sufficient to support a judgment of benefit,[43-49] the sample sizes were too small[44,47,48]; or there was a risk of conflicted interest.[45] Table 2 lists the Chinese herbal medicine reviews, and shows the number of significant meta-analyses, as well as the outcomes associated with these significant meta-analyses in cases where the authors' conclusions suggested benefit.

Table 2. Characteristics of Cochrane reviews of Chinese herbal medicine

Review title, author and date	Author conclusions*	Trials N	Participants N	Trials with allocation concealment N (%)	Forest Plots N	Meta-analyses N	Significant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Chinese herbal medicine for atopic eczema. Zhang et al., 2005[24]	A	4	159	0 (0)	0	0	N/A	N/A
Chinese herbal medicine for primary dysmenorrhoea. Zhu et al., 2008[50]	A	39	3475	2 (5)	31	16	15 (94)	Reduction of pain and improvement in overall symptoms
Chinese herbal medicine for schizophrenia. Rathbone et al., 2005[51]	A	7	1094	0 (0)	25	13	13 (100)	Improvement in global symptoms, negative symptoms
Chinese herbal medicines for the treatment of preeclampsia. Zhang, Wu, and Liu, 2006[52]	B	0	0	N/A	N/A	N/A	N/A	N/A
Chinese herbal medicine Huangqi type formulations for nephrotic syndrome. Yuan, Wang, and Wu, 2008[53]	A	3	128	0 (0)	11	0	N/A	N/A
Chinese herbal medicine suxiao jiuxin wan for angina pectoris. Duan et al., 2008[54]	A	15	1776	0 (0)	14	8	7 (88)	Electrocardiogram improvement and symptom improvement
Chinese herbal medicines for hyperthyroidism. Zen et al., 2007[55]	B	13	1770	0 (0)	55	0	N/A	N/A
Chinese herbal medicines for type 2 diabetes mellitus. Liu et al., 2004[56]	A	66	8302	0 (0)	146	3	3 (100)	Normalization of fasting blood glucose
Chinese herbal medicines in the treatment of ectopic pregnancy. Dengfeng et al., 2007[45]	B	2	157	0 (0)	15	2	2 (100)	N/A
Chinese herbs combined with Western medicine for severe acute respiratory syndrome (SARS). Liu et al., 2006[57]	A	13	654	0 (0)	37	2	0 (0)	N/A

Review title, author and date	Author conclu-sions*	Trials N	Partici-pants N	Trials with allocation conceal-ment N (%)	Forest Plots N	Meta-analyses N	Signifi-cant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Chinese medical herbs for chemotherapy side effects in colorectal cancer patients. Wu, Munro, and Guanjian, 2005[58]	B	4	342	0 (0)	5	5	4 (80)	N/A
Chinese medicinal herbs for acute bronchitis. Wei et al., 2008[59]	B	0	0	N/A	N/A	N/A	N/A	N/A
Chinese medicinal herbs for acute pancreatitis. Qiong et al., 2005[60]	A	11	658	0 (0)	0	0	N/A	N/A
Chinese medicinal herbs for asymptomatic carriers of hepatitis B virus infection. Liu, McIntosh, and Lin, 2001[61]	A	3	307	0 (0)	10	3	3 (100)	Loss of serum HBsAg, loss of serum HBeAg, and seroconversion of HBeAg to anti-HBe
Chinese medicinal herbs for chronic hepatitis B. Liu, McIntosh, and Lin, 2001[62]	A	9	936	1 (11)	78	15	11 (73)	Loss of serum HBsAg, loss of serum HBeAg, serocoversion of HBeAg to anti-HBe antibody, loss of serum HBV DNA, and serum alamine transaminase normalization
Chinese medicinal herbs for the common cold. Wu et al., 2007[63]	A	14	2440	4 (29)	23	4	0 (0)	N/A
Chinese medicinal herbs for influenza. Chen, et al., 2007[64]	B	2	1012	1 (50)	7	0	N/A	N/A
Chinese medicinal herbs for measles. Gu et al., 2006[65]	B	0	0	N/A	N/A	N/A	N/A	N/A
Chinese medicinal herbs for sore throat. Shi et al., 2007[66]	B	7	1253	0 (0)	21	0	N/A	N/A
Chinese medicinal herbs to treat the side-effects of chemotherapy in breast cancer patients. Zhang et al., 2007[67]	A	7	542	0 (0)	28	0	N/A	N/A

Review title, author and date	Author conclu-sions*	Trials N	Partici-pants N	Trials with allocation conceal-ment N (%)	Forest Plots N	Meta-analyses N	Signifi-cant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Chuanxiong-type preparations for acute ischemic stroke. Yuan et al., 2008[48]	B	2	161	0 (0)	3	1	1 (100)	N/A
Dan Shen agents for acute ischaemic stroke. Wu, Liu, and Zhang, 2007[46]	B	6	494	0 (0)	9	7	6 (86)	N/A
Danshen (Chinese medicinal herb) preparations for acute myocardial infarction. Wu, Ni, and Wu, 2008[47]	B	6	2368	0 (0)	11	1	1 (100)	N/A
Dengzhanhua preparations for acute cerebral infarction. Cao, 2008[44]	B	9	723	0 (0)	1	1	1 (100)	N/A
Elemene for the treatment of lung cancer. Rui et al., 2007[68]	B	0	0	N/A	N/A	N/A	N/A	N/A
Ginkgo biloba extract for age-related macular degeneration. Evans, 2000[69]	B	2	119	1 (50)	4	0	0 (0)	N/A
Ginkgo biloba for acute ischaemic stroke. Zeng et al., 2005[49]	B	10	792	1 (10)	5	4	4 (100)	N/A
Ginkgo biloba for cognitive impairment and dementia. Birks and Grimley Evans, 2007[43]	B	35	4291	17 (49)	64	34	12 (35)	N/A
Ginkgo biloba for tinnitus. Hilton and Stuart, 2004[70]	B	3	1143	3 (100)	0	N/A	N/A	N/A
Herbal medicines for treating HIV infection and AIDS. Liu, Manheimer, and Yang, 2005[71]	B	9	499	3 (33)	11	0	N/A	N/A
Herbal medicines for treatment of irritable bowel syndrome. Lim et al., 2006[37]	A	75	7957	3 (4)	112	0	N/A	N/A

Review title, author and date	Author conclu-sions*	Trials N	Partici-pants N	Trials with allocation conceal-ment N (%)	Forest Plots N	Meta-analyses N	Signifi-cant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Herbal medicines for viral myocarditis. Liu, Yang and Du, 2004[72]	A	40	3448	0 (0)	106	7	2 (29)	Improved arrhythmia and cardiac function and reduced myocardial enzyme lactate dehydrogenase levels
Huperzine A for Alzheimer's disease. Li, et al., 2008[73]	A	6	454	1 (17)	27	7	2 (29)	Improvement in cognitive function (mini-mental state examination) and functional performance (activities of daily living)
Medicinal herbs for hepatitis C virus infection. Liu et al., 2001[74]	B	10	517	2 (20)	28	0	N/A	N/A
Puerarin for acute ischaemic stroke. Tan, Liu, and Wu, 2008[75]	B	1	98	0 (0)	1	0	N/A	N/A
Puerarin injection for unstable angina pectoris. Wang, et al., 2006[76]	A	20	1240	0 (0)	8	5	3 (60)	Electrocardiogram improvements, reduction in angina attacks, and reduction in nitroglycerine daily dose
Sanchi for acute ischaemic stroke. Chen et al., 2008[77]	A	8	660	3 (38)	24	114	12 (86)	Improvement of neurological deficit, death or dependence at final follow-up
Shengmai (a traditional Chinese herbal medicine) for heart failure. Chen et al., 2007[78]	A	19	1663	0 (0)	22	1	1 (100)	Classification of clinical status; mortality; tumor necrosis factor-alpha
Tongxinluo capsule for acute stroke. Zhuo et al., 2008[79]	B	2	232	0 (0)	1	0	N/A	N/A
Tongxinluo (Tong xin luo or Tong-xin-luo) capsule for unstable angina pectoris. Wu et al., 2006[80]	A	18	1413	1 (6)	29	15	9 (60)	Improved electrocardiogram, reduced frequency of acute angina attack, improved angina symptoms
Yizhi capsule for vascular dementia. Wu, Li and Yuan, 2007[81]	B	0	0	N/A	N/A	N/A	N/A	N/A

Review title, author and date	Author conclusions*	Trials N	Participants N	Trials with allocation concealment N (%)	Forest Plots N	Meta-analyses N	Significant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Zhiling decoction for vascular dementia. Jirong et al., 2004[82]	B	0	0	N/A	N/A	N/A	N/A	N/A

*A = abstract conclusions statement indicated a suggestion of benefit, which was qualified by a caveat about the poor quality and quantity of included studies, and B = abstract conclusions statement indicated that the currently available data do not allow any conclusions to be drawn

†Outcomes are reported only for reviews which 1) include a significant meta-analysis, and 2) for which the authors' abstract conclusions statement indicated a suggestion of benefit

Other TCM therapies

We found two Cochrane reviews on TCM that were not acupuncture or herbal medicine. One review was on the use of moxibustion for cephalic version in breech presentation and one review was on Tai Chi in the treatment of rheumatoid arthritis. Both reviews identified trials, but only one review conducted meta-analyses. The author conclusion statements in these reviews were cautiously positive, and thus the reviews were placed in category A. These reviews are listed in Table 3.

Table 3. Characteristics of Cochrane reviews of Moxibustion and Tai Chi

Review title, author and date	Author conclusions*	Trials N	Participants N	Trials with allocation concealment N (%)	Forest Plots N	Meta-analyses N	Significant meta-analyses N (%)	Significant meta-analytic outcomes in reviews suggesting intervention benefits†
Cephalic version by moxibustion for breech presentation. Coyle et al., 2005[83]	A	3	597	2 (67)	12	0	N/A	N/A
Tai chi for treating rheumatoid arthritis. Han et al., 2004[84]	A	4	206	0 (0)	16	6	1 (17)	Withdrawals overall‡

*A = abstract conclusions statement indicated a suggestion of benefit, which was qualified by a caveat about the poor quality and quantity of included studies, and B = abstract conclusions statement indicated that the currently available data do not allow any conclusions to be drawn

†Outcomes are reported only for reviews which 1) include a significant meta-analysis, and 2) for which the authors' abstract conclusions statement indicated a suggestion of benefit

‡Findings of benefit of treatment were based upon improvements in lower limb flexion identified in a single trial and therefore no meta-analysis on this outcome was conducted. Fewer withdrawals from treatment in the tai chi group were reported by authors to be possibly due to greater enjoyment and perceived benefit by tai chi participants.

DISCUSSION

TCM therapies, including acupuncture and herbal medicine, are used increasingly to complement conventional medical care. Scientific research has grown in response to this surge in popular use, with the number of projects funded to study TCM approaches seeing a dramatic rise in the last decade. The Cochrane Collaboration has summarized much of the evidence on acupuncture and Chinese herbal medicine, and its *Cochrane Database of Systematic Reviews* currently includes 26

systematic reviews on acupuncture and 42 systematic reviews on Chinese herbal medicine. Many of these systematic reviews are inconclusive, due specifically to the poor methodology and heterogeneity of the studies reviewed. However, several systematic reviews provide preliminary evidence of Chinese medicine's benefits to certain patient populations. For example, acupuncture may be helpful for those with post-operative nausea and vomiting, chemotherapy-induced nausea and vomiting, various types of chronic pain, and as an adjuvant treatment for *in vitro* fertilization. Chinese herbal medicine may also be helpful for a wide range of conditions, including Huperzine A for Alzheimer's disease and Sanchi for acute ischemic stroke.

We indicated which reviews showed the most promising evidence by categorizing the authors' abstract conclusions statements, and for those statements suggesting a possible benefit, we listed the comparisons/outcomes that were statistically significant. An advantage of selecting the most promising reviews based on the abstract conclusions statements are that these statements provide the most informative and concise summary of the treatments' effects. Their drawback is that they do not include quantitative estimates of effect, and also they may be somewhat subjective or inaccurate, depending on the reviewer's interpretation of the evidence. Therefore, we supplemented the categorization of the abstract conclusions with a listing of the comparisons/outcomes showing statistically significant meta-analyses results.

However, benefits suggested in the meta-analyses need to be interpreted with much caution. The data generally have been pooled from trials with a variety of methodological quality, some of which were imprecise and carried a high risk of bias. Some reviews with positive abstract conclusion statements do not include any meta-analyses, due to the heterogeneity of the trials included, rendering the summary less quantitative and more qualitative. An additional limitation is that a majority of trials included here have been carried out and published in China. The quality of these trials typically has been low,[85,86] as indicated by non-compliance with CONSORT statement recommendations[87] such as describing randomization, allocation concealment, sample size calculation and participant follow-up procedures. The Chinese government recently has made substantial investments into funding Chinese medicine clinical research,[88] which already has resulted in important improvements and greater compliance to international standards.[89] Once a substantial number of trials of improved methodological quality are completed, future updates of currently inconclusive Cochrane reviews of Chinese medicine may reach more definitive conclusions, and true evidence-based Chinese medicine will be possible.

In summary, the current Cochrane systematic review evidence for TCM herbs is somewhat inconsistent and not yet fully convincing because of the methodological problems encountered when summarizing studies that are heterogeneous on the type of herbal formulas, use of control conditions, formula dosages and durations of use.[35] Rather than proclaiming the success of TCM approaches, our intention here is to highlight those trials that suggest possible benefit in order to encourage further rigorous study in these areas. These potential benefits need to be confirmed in large, rigorous, double-blind randomized placebo-controlled trials.

The reason we have restricted this overview to Cochrane reviews is because they are implicitly trustworthy, given the rigorous methodological standards imposed by the Cochrane Collaboration in its acceptance of articles. Indeed, Cochrane reviews require the use of explicit and transparent methods, are peer-reviewed at both the protocol and complete review stage, and are regularly updated. For these reasons, they have been found to be of comparable or better quality than reviews published in even the leading print journals.[90,91] We also restricted to Cochrane reviews because they use fairly homogeneous methods for both their preparation and their peer-review, as described in the Cochrane Reviewer's Handbook, and therefore across Cochrane reviews, the authors' conclusions statements for a given level of evidence would be fairly standard and consistent. If we had included non-Cochrane reviews, we may have introduced a source of

heterogeneity to the conclusions presented in the tables, due not to differences in the evidence from the RCTs reviewed but rather to a difference in the location of publication. A limitation of restricting to Cochrane reviews is that the breadth of TCM topics covered by the Cochrane Database is not yet comprehensive, and there are some topics that are currently still at the protocol or registered title stage. Two notable omissions from the tables are the absence of Cochrane reviews of acupuncture for post-operative pain and acupuncture for osteoarthritis; both of these reviews are currently published only as protocols in the *Cochrane Database*, but both should be available as full reviews within the next year. We are currently conducting a project to identify high priority but not yet registered Cochrane titles by comparing the lists of current Cochrane titles against the contents of the Cochrane Complementary Medicine Field register of trials to identify condition/treatment pairings for which RCTs have been published but no Cochrane title has yet been registered. The results of this project may identify other omissions in the current list of Cochrane reviews. A final limitation of our restricting to Cochrane reviews is that many of these reviews are several years old and their conclusions may not reflect the most current evidence. Although Cochrane policy states that reviews should ideally be updated every two years, in practice this is difficult to achieve because updates can involve a substantial amount of work, often approaching the effort level required to prepare a new review (which has been estimated at 1139 person hours[92]), and yet updates garner little academic recognition for the review authors. To address this, there are various initiatives underway to increase the frequency of updating of Cochrane reviews, including the provision of funding support from the US National Center for Complementary and Alternative Medicine (as disbursed through the Cochrane Complementary Medicine Field (http://medschool.umaryland.edu/integrative/cochrane_bursary.asp) and the UK Department of Health. Despite this limitation in updating frequency, Cochrane reviews are often still more current than reviews published in print journals, which are fixed and therefore may become obsolete shortly after publication.

CONCLUSIONS

For those seeking hard and fast answers to their healthcare questions, it may seem discouraging that the evidence for many therapies is undetermined. However, with the continued investment by federal funders such as the National Center for Complementary and Alternative Medicine at the National Institutes of Health, the Cochrane Collaboration and other organizations in supporting rigorous evaluation of TCM therapies, the evidence base will continue to grow. Questions that cannot be answered definitively today will become resolved in the future as current Cochrane reviews are updated to include newer and better trials. Twenty years ago, there were very few RCTs conducted on these modalities, and systematic review methodology was in its infancy. Now, after decades of work and innovation, true evidence-based traditional Chinese medicine is becoming a reality.

REFERENCES

1. Barnes PM, Powell-Griner E, McFann K, Nahin RL. Complementary and alternative medicine use among adults: United States, 2002. *Adv Data* 2004(343):1-19.
2. Wu AP, Burke A, LeBaron S. Use of traditional medicine by immigrant Chinese patients. *Fam Med* 2007;39(3):195-200.
3. Tindle HA, Davis RB, Phillips RS, Eisenberg DM. Trends in use of complementary and alternative medicine by US adults: 1997-2002. *Altern Ther Health Med* 2005;11(1):42-9.
4. Burke A, Upchurch DM, Dye C, Chyu L. Acupuncture use in the United States: findings from the National Health Interview Survey. *J Altern Complement Med* 2006;12(7):639-48.
5. Astin JA. Why patients use alternative medicine: results of a national study. *Jama* 1998;279(19):1548-53.
6. Lai D, Chappell N. Use of Traditional Chinese Medicine by older Chinese immigrants in Canada. *Fam Pract* 2007;24(1):56-64.
7. Chen FP, Chen TJ, Kung YY, Chen YC, Chou LF, Chen FJ, Hwang SJ. Use frequency of traditional Chinese medicine in Taiwan. *BMC Health Serv Res* 2007;7:26.
8. Chung V, Wong E, Woo J, Lo SV, Griffiths S. Use of traditional chinese medicine in the Hong Kong special administrative region of China. *J Altern Complement Med* 2007;13(3):361-7.
9. Cassidy CM. Chinese medicine users in the United States. Part I: Utilization, satisfaction, medical plurality. *J Altern Complement Med* 1998;4(1):17-27.
10. Kim J, Chan MM. Factors influencing preferences for alternative medicine by Korean Americans. *Am J Chin Med* 2004;32(2):321-9.
11. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M, Kessler RC. Trends in alternative medicine use in the United States, 1990-1997: results of a follow-up national survey. *Jama* 1998;280(18):1569-75.
12. Dickersin K, Manheimer E. The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin Obstet Gynecol* 1998;41(2):315-31.
13. Allen C, Clarke M, Tharyan P. International activity in the Cochrane Collaboration with particular reference to India. *Natl Med J India* 2007;20(5):250-5.
14. Commercial Sponsorship and the Cochrane Collaboration (updated April 2006). Available at: http://www.cochrane.org/about-us/commercial-sponsorship (Accessed 2013 June 12).
15. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336(7644):601-5.
16. Melchart D, Linde K, Fischer P, Berman B, White A, Vickers A, Allais G. Acupuncture for idiopathic headache. *Cochrane Database Syst Rev* 2001(1):CD001218.
17. Trinh KV, Graham N, Gross AR, Goldsmith CH, Wang E, Cameron ID, Kay T. Acupuncture for neck disorders. *Cochrane Database Syst Rev* 2006;3:CD004870.
18. Cheuk DK, Wong V. Acupuncture for epilepsy. *Cochrane Database Syst Rev* 2008(4):CD005062.
19. Cheuk DK, Yeung WF, Chung KF, Wong V. Acupuncture for insomnia. *Cochrane Database Syst Rev* 2007(3):CD005472.
20. Rathbone J, Xia J. Acupuncture for schizophrenia. *Cochrane Database Syst Rev* 2005(4):CD005475.
21. Smith CA, Crowther CA. Acupuncture for induction of labour. *Cochrane Database Syst Rev* 2004(1):CD002962.
22. White AR, Rampes H, Campbell JL. Acupuncture and related interventions for smoking cessation. *Cochrane Database Syst Rev* 2006(1):CD000009.
23. Wu HM, Tang JL, Lin XP, Lau J, Leung PC, Woo J, Li YP. Acupuncture for stroke rehabilitation. *Cochrane Database Syst Rev* 2006;3:CD004131.
24. Zhang W, Leonard T, Bath-Hextall F, Chambers CA, Lee C, Humphreys R, Williams HC. Chinese herbal medicine for atopic eczema. *Cochrane Database Syst Rev* 2005(2):CD002291.
25. Zhang SH, Liu M, Asplund K, Li L. Acupuncture for acute stroke. *Cochrane Database Syst Rev* 2005(2):CD003317.
26. Cheong YC, Hung Yu Ng E, Ledger WL. Acupuncture and assisted conception. *Cochrane Database Syst Rev* 2008(4):CD006920.
27. Casimiro L, Barnsley L, Brosseau L, Milne S, Robinson VA, Tugwell P, Wells G. Acupuncture and electroacupuncture for the treatment of rheumatoid arthritis. *Cochrane Database Syst Rev* 2005(4):CD003788.
28. Furlan AD, van Tulder MW, Cherkin DC, Tsukayama H, Lao L, Koes BW, Berman BM. Acupuncture and dry-needling for low back pain. *Cochrane Database Syst Rev* 2005(1):CD001351.
29. He L, Zhou MK, Zhou D, Wu B, Li N, Kong SY, Zhang DP, Li QF, Yang J, Zhang X. Acupuncture for Bell's palsy. *Cochrane Database Syst Rev* 2007(4):CD002914.
30. McCarney RW, Brinkhaus B, Lasserson TJ, Linde K. Acupuncture for chronic asthma. *Cochrane Database Syst Rev* 2004(1):CD000008.
31. Smith CA, Hay PP. Acupuncture for depression. *Cochrane Database Syst Rev* 2005(2):CD004046.
32. Xie Y, Wang L, He J, Wu T. Acupuncture for dysphagia in acute stroke. *Cochrane Database Syst Rev* 2008(3):CD006076.

33. Law SK, Li T. Acupuncture for glaucoma. *Cochrane Database Syst Rev* 2007(4):CD006030.
34. Green S, Buchbinder R, Barnsley L, Hall S, White M, Smidt N, Assendelft W. Acupuncture for lateral elbow pain. *Cochrane Database Syst Rev* 2002(1):CD003527.
35. Cui Y, Wang Y, Liu Z. Acupuncture for restless legs syndrome. *Cochrane Database Syst Rev* 2008(4):CD006457.
36. Green S, Buchbinder R, Hetrick S. Acupuncture for shoulder pain. *Cochrane Database Syst Rev* 2005(2):CD005319.
37. Lim B, Manheimer E, Lao L, Ziea E, Wisniewski J, Liu J, Berman B. Acupuncture for treatment of irritable bowel syndrome. *Cochrane Database Syst Rev* 2006(4):CD005111.
38. Peng WN, Zhao H, Liu ZS, Wang S. Acupuncture for vascular dementia. *Cochrane Database Syst Rev* 2007(2):CD004987.
39. Ezzo JM, Richardson MA, Vickers A, Allen C, Dibble SL, Issell BF, Lao L, Pearl M, Ramirez G, Roscoe J, et al. Acupuncture-point stimulation for chemotherapy-induced nausea or vomiting. *Cochrane Database Syst Rev* 2006(2):CD002285.
40. Gates S, Smith LA, Foxcroft DR. Auricular acupuncture for cocaine dependence. *Cochrane Database Syst Rev* 2006(1):CD005192.
41. Lee A, Done ML. Stimulation of the wrist acupuncture point P6 for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev* 2004(3):CD003281.
42. Proctor ML, Smith CA, Farquhar CM, Stones RW. Transcutaneous electrical nerve stimulation and acupuncture for primary dysmenorrhoea. *Cochrane Database Syst Rev* 2002(1):CD002123.
43. Birks J, Grimley Evans J. Ginkgo biloba for cognitive impairment and dementia. *Cochrane Database Syst Rev* 2007(2):CD003120.
44. Cao W, Liu W, Wu T, Zhong D, Liu G. Dengzhanhua preparations for acute cerebral infarction. *Cochrane Database Syst Rev* 2008(4):CD005568.
45. Dengfeng W, Taixiang W, Lina H, Marjoribanks J, Guanjian L, Haijun J, Ying S, Jing Z. Chinese herbal medicines in the treatment of ectopic pregnancy. *Cochrane Database Syst Rev* 2007(4):CD006224.
46. Wu B, Liu M, Zhang S. Dan Shen agents for acute ischaemic stroke. *Cochrane Database Syst Rev* 2007(2):CD004295.
47. Wu T, Ni J, Wu J. Danshen (Chinese medicinal herb) preparations for acute myocardial infarction. *Cochrane Database Syst Rev* 2008(2):CD004465.
48. Yuan Y, Zeng X, Luo Y, Li Z, Wu T. Chuanxiong-type preparations for acute ischemic stroke. *Cochrane Database Syst Rev* 2008(4):CD005569.
49. Zeng X, Liu M, Yang Y, Li Y, Asplund K. Ginkgo biloba for acute ischaemic stroke. *Cochrane Database Syst Rev* 2005(4):CD003691.
50. Zhu X, Proctor M, Bensoussan A, Wu E, Smith CA. Chinese herbal medicine for primary dysmenorrhoea. *Cochrane Database Syst Rev* 2008(2):CD005288.
51. Rathbone J, Zhang L, Zhang M, Xia J, Liu X, Yang Y. Chinese herbal medicine for schizophrenia. *Cochrane Database Syst Rev* 2005(4):CD003444.
52. Zhang J, Wu TX, Liu GJ. Chinese herbal medicine for the treatment of pre-eclampsia. *Cochrane Database Syst Rev* 2006(2):CD005126.
53. Yuan W, Wang J, Wu T. Chinese herbal medicine Huangqi type formulations for nephrotic syndrome. *Cochrane Database Syst Rev* 2008(2):CD006335.
54. Duan X, Zhou L, Wu T, Liu G, Qiao J, Wei J, Ni J, Zheng J, Chen X, Wang Q. Chinese herbal medicine suxiao jiuxin wan for angina pectoris. *Cochrane Database Syst Rev* 2008(1):CD004473.
55. Zen XX, Yuan Y, Liu Y, Wu TX, Han S. Chinese herbal medicines for hyperthyroidism. *Cochrane Database Syst Rev* 2007(2):CD005450.
56. Liu JP, Zhang M, Wang WY, Grimsgaard S. Chinese herbal medicines for type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2004(3):CD003642.
57. Liu X, Zhang M, He L, Li YP, Kang YK. Chinese herbs combined with Western medicine for severe acute respiratory syndrome (SARS). *Cochrane Database Syst Rev* 2006(1):CD004882.
58. Wu T, Munro AJ, Guanjian L. Chinese medical herbs for chemotherapy side effects in colorectal cancer patients. *Cochrane Database Syst Rev* 2005(1):CD004540.
59. Wei J, Ni J, Wu T, Chen X, Duan X, Liu G, Qiao J, Wang Q, Zheng J, Zhou L. Chinese medicinal herbs for acute bronchitis. *Cochrane Database Syst Rev* 2008(1):CD004560.
60. Qiong W, Yiping W, Jinlin Y, Tao G, Zhen G, Pengcheng Z. Chinese medicinal herbs for acute pancreatitis. *Cochrane Database Syst Rev* 2005(1):CD003631.
61. Liu JP, McIntosh H, Lin H. Chinese medicinal herbs for asymptomatic carriers of hepatitis B virus infection. *Cochrane Database Syst Rev* 2001(2):CD002231.
62. Liu JP, McIntosh H, Lin H. Chinese medicinal herbs for chronic hepatitis B. *Cochrane Database Syst Rev* 2001(1):CD001940.

63. Wu T, Zhang J, Qiu Y, Xie L, Liu GJ. Chinese medicinal herbs for the common cold. *Cochrane Database Syst Rev* 2007(1):CD004782.
64. Chen XY, Wu TX, Liu GJ, Wang Q, Zheng J, Wei J, Ni J, Zhou LK, Duan X, Qiao JQ. Chinese medicinal herbs for influenza. *Cochrane Database Syst Rev* 2007(4):CD004559.
65. Gu R, Shi YY, Wu TX, Liu GJ, Zhang MM. Chinese medicinal herbs for measles. *Cochrane Database Syst Rev* 2006(2):CD005531.
66. Shi Y, Gu R, Liu C, Ni J, Wu T. Chinese medicinal herbs for sore throat. *Cochrane Database Syst Rev* 2007(3):CD004877.
67. Zhang M, Liu X, Li J, He L, Tripathy D. Chinese medicinal herbs to treat the side-effects of chemotherapy in breast cancer patients. *Cochrane Database Syst Rev* 2007(2):CD004921.
68. Rui D, Xiaoyan C, Taixiang W, Guanjian L. Elemene for the treatment of lung cancer. *Cochrane Database Syst Rev* 2007(4):CD006054.
69. Evans JR. Ginkgo biloba extract for age-related macular degeneration. *Cochrane Database Syst Rev* 2000(2):CD001775.
70. Hilton M, Stuart E. Ginkgo biloba for tinnitus. *Cochrane Database Syst Rev* 2004(2):CD003852.
71. Liu JP, Manheimer E, Yang M. Herbal medicines for treating HIV infection and AIDS. *Cochrane Database Syst Rev* 2005(3):CD003937.
72. Liu JP, Yang M, Du XM. Herbal medicines for viral myocarditis. *Cochrane Database Syst Rev* 2004(3):CD003711.
73. Li J, Wu HM, Zhou RL, Liu GJ, Dong BR. Huperzine A for Alzheimer's disease. *Cochrane Database Syst Rev* 2008(2):CD005592.
74. Liu JP, Manheimer E, Tsutani K, Gluud C. Medicinal herbs for hepatitis C virus infection. *Cochrane Database Syst Rev* 2001(4):CD003183.
75. Tan Y, Liu M, Wu B. Puerarin for acute ischaemic stroke. *Cochrane Database Syst Rev* 2008(1):CD004955.
76. Wang Q, Wu T, Chen X, Ni J, Duan X, Zheng J, Qiao J, Zhou L, Wei J. Puerarin injection for unstable angina pectoris. *Cochrane Database Syst Rev* 2006;3:CD004196.
77. Chen X, Zhou M, Li Q, Yang J, Zhang Y, Zhang D, Kong S, Zhou D, He L. Sanchi for acute ischaemic stroke. *Cochrane Database Syst Rev* 2008(4):CD006305.
78. Chen J, Wu G, Li S, Yu T, Xie Y, Zhou L, Wang L. Shengmai (a traditional Chinese herbal medicine) for heart failure. *Cochrane Database Syst Rev* 2007(4):CD005052.
79. Zhuo Q, Yang X, Wu T, Liu G, Zhou L. Tongxinluo capsule for acute stroke. *Cochrane Database Syst Rev* 2008(4):CD004584.
80. Wu T, Harrison RA, Chen X, Ni J, Zhou L, Qiao J, Wang Q, Wei J, Xin D, Zheng J. Tongxinluo (Tong xin luo or Tong-xin-luo) capsule for unstable angina pectoris. *Cochrane Database Syst Rev* 2006(4):CD004474.
81. Wu TX, Li QP, Yuan ZY. Yizhi capsule for vascular dementia. *Cochrane Database Syst Rev* 2007(2):CD005382.
82. Jirong Y, Xiaoyan Y, Taixiang W, Defen S, Birong D. Zhiling decoction for vascular dementia. *Cochrane Database Syst Rev* 2004(4):CD004670.
83. Coyle ME, Smith CA, Peat B. Cephalic version by moxibustion for breech presentation. *Cochrane Database Syst Rev* 2005(2):CD003928.
84. Han A, Robinson V, Judd M, Taixiang W, Wells G, Tugwell P. Tai chi for treating rheumatoid arthritis. *Cochrane Database Syst Rev* 2004(3):CD004849.
85. Tang JL, Zhan SY, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;319(7203):160-1.
86. Zhang D, Yin P, Freemantle N, Jordan R, Zhong N, Cheng K. An assessment of the quality of randomised controlled trials conducted in China. *Trials* 2008;9:22.
87. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134(8):657-62.
88. Qiu J. China plans to modernize traditional medicine. *Nature* 2007;446(7136):590-1.
89. Wang G, Mao B, Xiong ZY, Fan T, Chen XD, Wang L, Liu GJ, Liu J, Guo J, Chang J, et al. The quality of reporting of randomized controlled trials of traditional Chinese medicine: a survey of 13 randomly selected journals from mainland China. *Clin Ther* 2007;29(7):1456-67.
90. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280(3):278-80.
91. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4(3):e78.
92. Glasziou P, Irwig L, Bain C, Colditz G. *Systematic reviews in health care: a practical guide*. Cambridge: Cambridge University Press, 2001.

Chapter 5. Systematic reviews and CAM

Manheimer E*, Ezzo J*. Chapter 2: Systematic reviews and CAM. In: Adams J (editor). Researching Complementary and Alternative Medicine. Oxford: Routledge, 2006, pgs 19-36. *Both authors contributed equally to this work.

INTRODUCTION

Most systematic reviews of CAM restrict inclusion to RCTs, widely regarded as the most unbiased study design for evaluating health-care interventions. Systematic reviewers evaluate and synthesise RCTs using objective, transparent and reproducible methods in order to assess the overall effects of a given therapy and systematic reviews sometimes include a meta-analysis, the quantitative combining (pooling) of results from similar but separate RCTs to obtain an overall effect estimate.

Over the past twenty years, there has been an explosion in the number of meta-analyses in CAM (Figure 1). Meta-analysis now has the greatest citation impact of all study designs (exceeding even RCTs) and is continuing to increase.[1] This citation impact of meta-analysis/systematic review is also commensurate with its position at the top of the hierarchy of research evidence[2] and the recent interest in CAM from the Cochrane Collaboration[3] (as of July 2005, there were 2,435 completed Cochrane reviews and more than 150 CAM-related Cochrane reviews). This chapter provides an overview of systematic review methods in relation to CAM, summarises current research on CAM systematic reviews and illustrates through case examples various approaches used to address methodological challenges in CAM reviews.

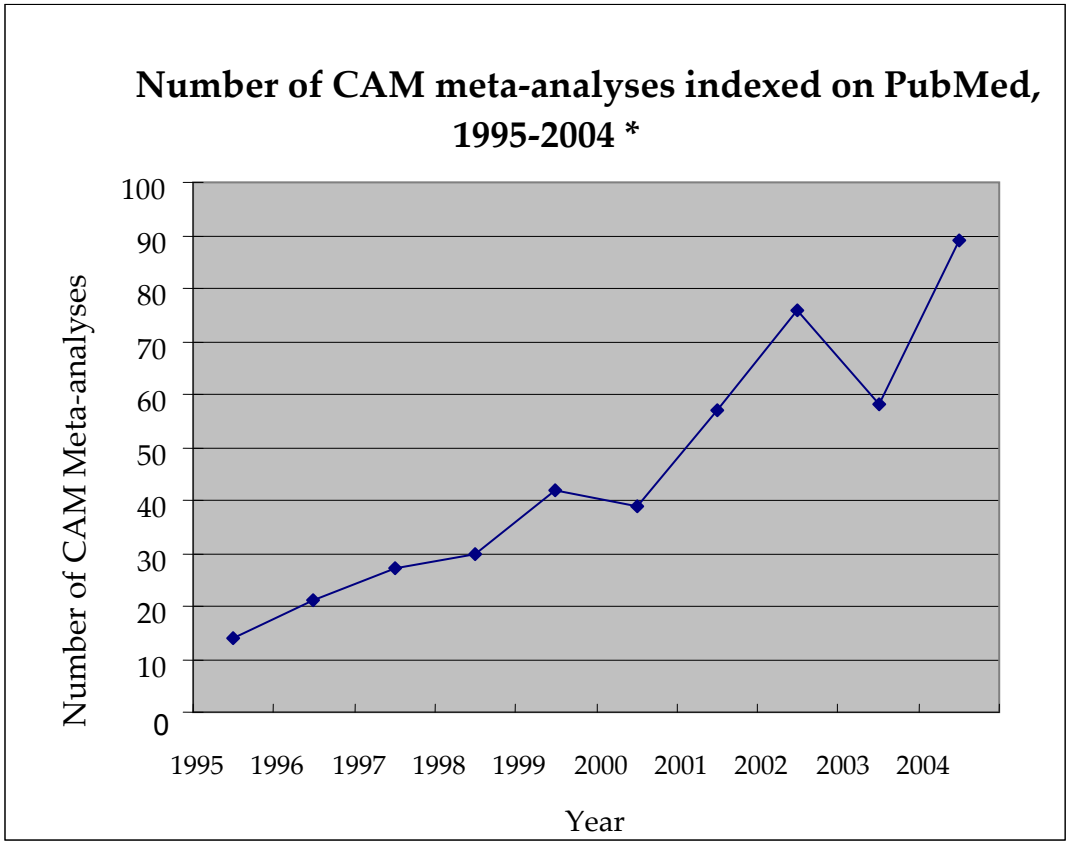


Figure 1. Number of CAM meta-analyses indexed on PubMed, 1995-2004

*This search was performed on October 14, 2005 using the following search strategy to obtain counts for each year: CAM [Subset] AND meta-analysis [Publication Type] AND *year* [Date of Publication]

Because there is no MEDLINE Publication Type term for *systematic review*, we used the Publication Type term *meta-analysis* as an indicator for tracking growth in interest in systematic reviews over the past ten years. While the term *meta-analysis* is likely to have a high precision in identifying systematic reviews (because a meta-analysis is generally also a systematic review), this term is likely to have only a low to moderate sensitivity (because many systematic reviews do not include a meta-analysis).

SYSTEMATIC REVIEWS: THEIR IMPORTANCE TO RESEARCH

Systematic reviews are rapidly becoming the cornerstone of evidence-based medicine with clinicians ranking reviews as the primary source of new information.[4] Policy-makers increasingly rely on systematic reviews as a way of summarising evidence[5] and consumers use reviews to guide health decisions.[6]

Information from systematic reviews also aid researchers in their attempts to plan clinical trials. The systematic review serves to ensure that the proposed trial is relevant, necessary and guided by earlier trials. Amidst the vast, almost limitless, number of research questions that remain to be addressed in CAM, and the limited financial support available to study non-proprietary CAM therapies, it is important that researchers plan their trials in the context of what is already known on a topic. Having spent months studying the existing CAM trials, systematic reviewers are well versed in the strengths and weaknesses of current trials and often ideally suited to suggest methodological improvements for future trials. For example, Berman et al.[7] designed a large, phase III 'acupuncture for knee osteoarthritis' trial using guidelines from an earlier review[8] and McNeely et al.[9] cite the methodological limitations of an earlier systematic review as a stimulus for their recent research design.

Systematic reviews have a two-way, iterative relationship with clinical trials, and this is well illustrated with the example of acupuncture for low back pain. The earliest Cochrane review of trials on this topic was inconclusive due to methodological weaknesses.[10] Larger, more rigorous trials were conducted, addressing the issues raised in the Cochrane review, and the two most recent systematic reviews of acupuncture for low back pain[11,12] conducted by two independent research teams both show positive findings favouring acupuncture compared to control for chronic low back pain.

Systematic reviews can also influence primary research by suggesting priorities for investigation. Some reviews now include suggestions for high priority research based on known mechanisms of action and safety. For example, a review of CAM for dementia cited huperzine A, levacecarnine, and EGB 761 as warranting further examination based on the methodological quality of the studies, mechanisms of action and overall safety.[13]

SYSTEMATIC REVIEWS REQUIRE RIGOROUS METHODS

Systematic reviews can be prone to the biases that also plague other research study designs. In the context of systematic reviews, the term 'bias' is used to designate some systematic study-related error resulting in the failure to reflect the real world association between treatment and outcome. The susceptibility to bias within systematic reviews is illustrated in a number of ways. First, systematic review findings have occasionally been overturned by the findings of large, well-designed RCTs.[14] Second, as Linde and Willich[15] illustrate with regard to acupuncture, herbal medicine and homoeopathy, systematic reviews that address the same research question sometimes employ different methods of review leading to differences in results and conclusions. Systematic reviews are designed to ensure rigorous quality standards and maintain objectivity during each phase of review preparation, including: (1) identifying relevant RCTs; (2) assessing the quality of the RCTs; and (3) combining the data from the RCTs. Issues regarding these three items are discussed in the following three sections.

Identifying relevant trials

Conducting a thorough, well-documented search for trials is one of the key elements that distinguishes a systematic review from a traditional narrative review. While comprehensive searches of multiple-database and non-database sources of all languages are ideal under optimal circumstances, such far-reaching searches are not always practical given time and budget constraints. As a result, thoroughness needs to be balanced with efficiency. The best way to achieve this balance is to be aware of, and aim to minimise, the various biases that can result from restricting searches in different ways.

Can searches be restricted to major databases?

Research examining searches restricted to only the US National Library of Medicine's Medline database (or Medline and other major databases) shows these methods yield non-comprehensive results. Medline sensitivity averaged 51 per cent (range 17-82 per cent), for a sample of studies including both CAM and conventional medicine topics, even when databases were searched by a trained searcher.[16]

When searches are restricted to only trials in journals available on Medline, a substantial proportion (23 per cent) remain unidentified due to the inconsistent terminology employed to index randomised trials.[16] Such indexing difficulties have been documented for acupuncture trials[17] and CAM trials more generally.[18]

More current research on locating trials is pertinent given that indexing, coverage of databases and trial reporting have all improved in recent years.[19] A particular problem regarding CAM RCTs, which are often published in low-impact journals that are not a high indexing priority, is indexing lag time. For example, while extensive searches beyond the major databases have been shown to be necessary to identify nine of the twenty-one RCTs included in a systematic review of acupuncture,[20] post-hoc analysis shows that most of these RCTs have been indexed in the major databases one year later.

Other CAM researchers, conducting a systematic review of nutritional dietary supplements for patients after hip fracture, have evaluated the yield of RCTs by supplementing major database searches with other searches (for example, contacting experts, hand searching journals).[21] In this case, a search of only Medline and Embase articles would have missed approximately half of the eligible RCTs for this review. This is not surprising because Medline and Embase often exclude journals (often published in certain countries or languages[22]) which are likely to report CAM trials. Indeed, the proportion of non-Medline-indexed trials in CAM related meta-analyses (40.9 per cent) is approximately twice the proportion in conventional medicine meta-analyses (22.4 per cent).[23]

Can reviews exclude non-English-language publications?

Research has also examined the impact of including as opposed to excluding non-English-language trials. Such an impact is important because the identification and translation of non-English-language trials will substantially add to the costs of a review and will require the involvement of an international review team.

Several studies have shown, to a greater or lesser degree, that excluding trials published in non-English languages does not appear to substantially change effect estimates in meta-analyses of conventional medicine.[23-26] Meanwhile, Pham et al.[27] have shown that excluding non-English-language trials in CAM meta-analyses does change the effect estimates. The picture remains inconclusive and, ultimately, more research needs to be done to determine whether there are differential publication trends for CAM according to language, country and CAM modality.

Given the lack of a generalisable conclusion about non-English languages and CAM, there is fairly widespread agreement that for CAM, where a substantial proportion of the studies are not

included in Medline and other easy to access sources,[22] a non-comprehensive search may miss many eligible trials.[23] Although the studies that prove difficult to retrieve may be of lower quality,[23] it has been suggested that the correct approach is to not exclude them, but rather to evaluate their effects on the results of the review using sensitivity analyses.[23,26]

Assessing the methodological quality of RCTs

RCTs are the gold standard for evaluating the effects of health care therapies. However, the quality of RCTs is not uniform. Lower quality RCTs result in larger, and presumably inflated, effect estimates compared with RCTs of higher quality.[23,28] As a result, evaluating RCT quality has become a standard component of systematic reviews methodology[29] and this is as important for CAM as for any other area of health care research. We do not provide a detailed discussion of this area here (For a detailed discussion of RCTs, their applicability and quality with regard to CAM see Chapter 4 in this *Researching Complementary and Alternative Medicine* collection). However, one issue we would like to contemplate is the reporting of quality in RCT publications.

Quality is generally evaluated based on the information from the RCT publication, thus assuming that what is written in the publication reflects actual study procedure and that 'if it is not reported, it probably was not done.' In conventional medicine, existing studies covering different topic areas and publication periods have had contradictory findings on the utility of contacting investigators to obtain additional, unreported information about trial quality. A summary of the studies is published elsewhere.[30]

We recently evaluated this question in CAM by contacting principal investigators of acupuncture RCTs to request information about randomization and blinding procedures not described in RCT publications.[30] The investigation identified that over one-third of the trials had used appropriate random allocation concealment methods, but the investigators had failed to describe the details in their publications. While this survey suggests that contacting CAM trialists may result in obtaining previously unpublished information about methodological quality, the potential gains of obtaining the missing information may be outweighed by the reporting bias such efforts may introduce; data obtained directly from investigators has not been peer-reviewed and may not be as reliable as data extracted from published articles. Inadequate trial reporting is becoming less of a problem as a result of the CONSORT statement[19] providing a set of guidelines specifying reporting requirements for RCTs. An adaptation of CONSORT specifically for acupuncture trials, called STRICTA[31] has been widely disseminated in CAM journals.

Combining the data from RCTs in a meta-analysis: a case study

In deciding whether and how to statistically pool the results of similar but separate RCTs, systematic reviewers must consider the homogeneity of the populations studied, the therapies administered and the control comparisons used, as well as the homogeneity of the design and results of the RCTs. This is well illustrated with the use of a recently conducted systematic review and meta-analysis of acupuncture for low back pain.[12]

Heterogeneity of the trials was an important concern for this systematic review because the effects of acupuncture may vary depending on the style of acupuncture evaluated (Chinese or Western), the type of control comparison (sham, no treatment or another active treatment) and the type of pain in the patients studied (acute or chronic). To address potential heterogeneity, we decided, a priori, that eligible RCTs would be pooled in a meta-analysis only if they tested the same style of acupuncture against the same type of control for patients with the same type of low back pain. It transpired that a majority (twenty-two out of thirty-three) of the RCTs eligible for the systematic review evaluated Chinese-style acupuncture for patients with chronic low back pain.

The results of these RCTs, which were generally of fairly high quality, were meta-analysed together and stratified by control group in the primary analysis.

Figure 2 shows the results of this primary analysis, as a forest plot, the standard diagram for presenting meta-analysis results. The structure of the forest plot and the significance of the placement of the horizontal lines will be explained using the example of the sham-acupuncture controlled RCTs, grouped together at the top of the diagram. The horizontal lines associated with three out of four of these sham-acupuncture controlled RCTs did not cross the central vertical line of no effect, which indicates that these three trials all found acupuncture to be statistically significantly better than the sham-acupuncture control. The pooled result of all four of the sham-acupuncture controlled RCTs is indicated by the open circle (signifying the effect estimate) intersected by a horizontal line (signifying the confidence interval of this effect estimate).

The diagram indicates that both the sham controlled RCTs and the no-treatment controlled RCTs suggest acupuncture to be an effective treatment for relieving pain in the short term. However, the sham controlled RCTs generally show less benefit of acupuncture compared with the no-treatment controlled RCTs. The less positive outcome in the sham controlled RCTs is not surprising considering the potential for sham controlled RCTs to underestimate the specific effects of acupuncture (especially if the sham needles are inserted, thereby potentially stimulating a physiologic response)[32] and for no-treatment controlled trials (which are not blinded) to overestimate acupuncture's specific effects.

While the similarities of the RCTs' clinical characteristics, as described above, seemed to justify their pooling across control groups, before actually pooling the data, we also considered the separate but related issue of whether similarity in the statistical results of the RCTs could justify their statistical pooling. If the sham-acupuncture controlled RCTs had wildly different results, for example, then one might question the appropriateness of statistical pooling, even if the clinical characteristics of these trials seemed similar. Such pooling would especially be a concern if a difference in results were associated with a difference in design or methodological quality between the trials, for example. In this meta-analysis, we used statistical tests[33] to examine whether the results of the effects of acupuncture versus sham acupuncture were heterogeneous at different levels of any quality (for example, concealed allocation or not), patient (severe pain or not), or treatment-related (for example, number of sessions) criteria. We found that the results of these sham-acupuncture controlled trials were clearly homogeneous (see Figure 2) and no results on heterogeneity tests were positive for any criteria tested. The homogeneity of results (as indicated by the fact that the horizontal lines associated with the sham controlled trials all overlap) strengthened our confidence in both the appropriateness of using meta-analysis in this review and in the results of the review.

As mentioned above, twenty-two of the thirty-three studies in this systematic review evaluated the same style of acupuncture among patients with the same type of pain and were therefore judged sufficiently homogeneous to pool in our meta-analyses (see Figure 2). For the remaining eleven RCTs, which were fundamentally heterogeneous on style of acupuncture, type of pain and control comparison used, we employed a narrative description along with a tabular presentation of study characteristics and results instead of a meta-analysis. We did not meta-analyse any subset of these eleven trials due to the small number of RCTs within each subset, the small sample sizes and often poor reporting and low quality.

CAM systematic reviewers often cannot conduct a meta-analysis because of a dearth of available RCTs or deficiencies in the conduct and reporting of existing RCTs. A best-evidence synthesis is a qualitative alternative to a strictly narrative approach.[34] This method evaluates the consistency, quality and strength of the reviewed RCTs and, based on this evaluation, assigns a therapy a level of evidence: strong, moderate, limited, conflicting or none.

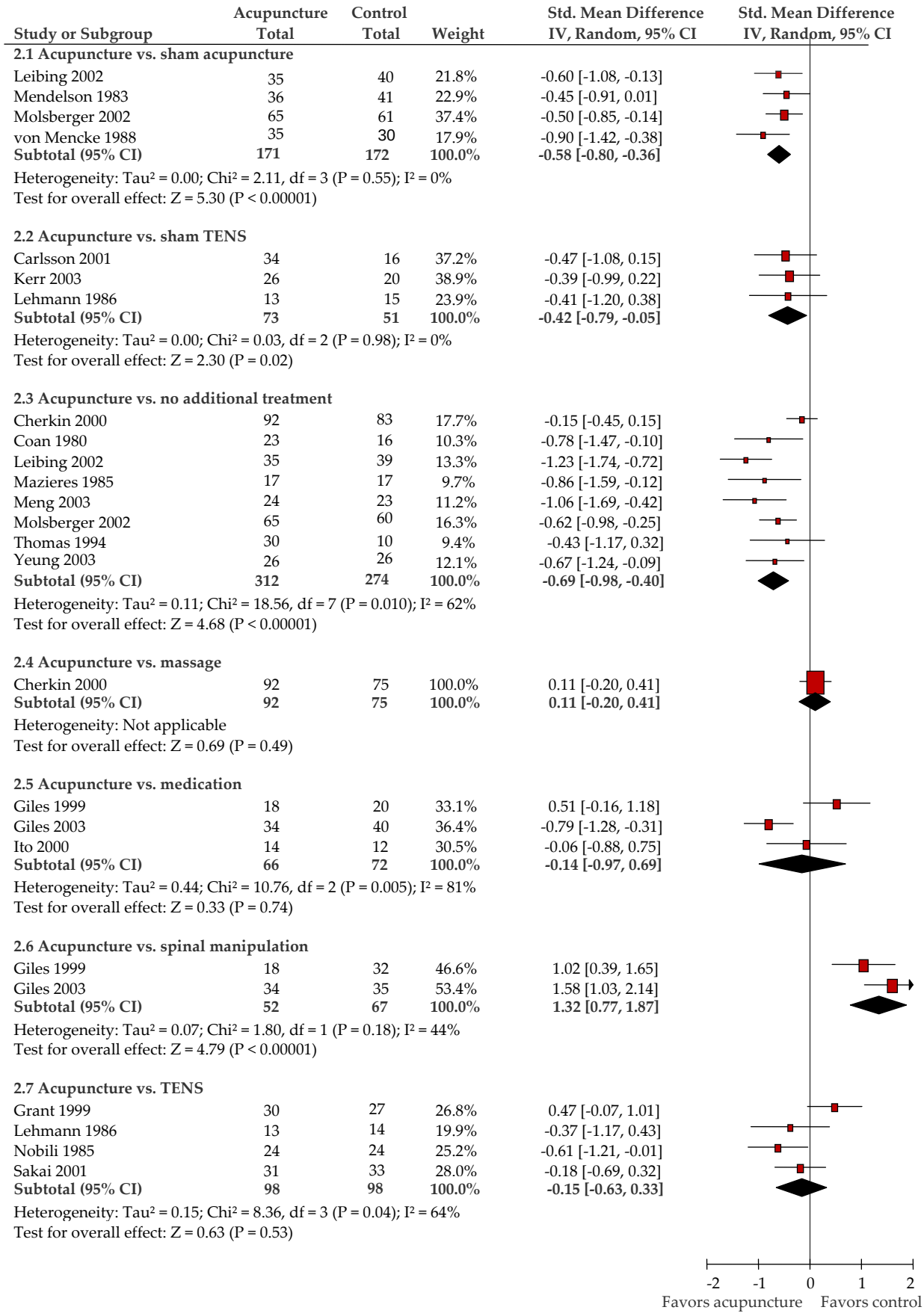


Figure 2. Meta-analysis forest plot: short-term effects of Chinese-style acupuncture on chronic pain
Adapted from original source: Figure 1 from Manheimer E, et al. Meta-analysis: acupuncture for low back pain. *Ann Intern Med*. 2005;142:651-663. Erratum in: *Ann Intern Med*. 2005 Jun 7;142(11):950-1.

ADDITIONAL RESEARCH ISSUES IN CAM

CAM reviewers commonly encounter additional methodological challenges. These are often judgment calls -situations where there is not one right answer but where reviewers have to choose one method over another, knowing the imperfections of each. Some common methodological issues include how to address treatment adequacy, practitioner adequacy, cointerventions and safety.

Assessment of the treatment adequacy

The discussion about trial validity in the first part of this chapter pertains to design issues such as concealment and blinding that are common across trials regardless of whether the trial is a drug or a CAM trial. Failure to address these issues can bias the review conclusions predominantly on the side of a type I error (a false-positive finding). However, many CAM interventions lack the dose-finding Phase I and II research of drug trials, and inadequate 'doses' can bias results towards a type II error (a false-negative finding).

Although treatment adequacy should be addressed in clinical trials, systematic reviews need a method for assessing it. Acupuncture reviews provide examples of a variety of ways treatment adequacy has been addressed. One treatment-adequacy assessment first proposed by Linde et al.[35] involves presenting inclusion criteria and methods sections of acupuncture papers to acupuncturists. These acupuncturists are blinded to the trial results and asked to rate whether the acupuncture provided was adequate to address the condition based on five aspects of treatment (the points selected, the total number of treatments, the number of times per week the patient was treated, the duration of each session and whether or not *de qi* was elicited). Unfortunately, this method proved very complex and the data were not interpretable (personal communication).

A second approach used by Molsberger and Bowing[36] defines a minimally adequate acupuncture treatment as consisting of at least ten total treatments of at least fifteen minutes each and a description of the points used. Only sixteen of eighty-eight studies on musculoskeletal or neurological conditions met these minimal criteria.

A third approach has used textbooks from China, Japan and Korea[37] to formulate criteria for treatment adequacy.[38] This work tests the hypotheses that: six acupuncture points selected per treatment session are adequate but ten points are even better, six total treatment sessions are adequate but ten sessions are even better, and that both of these parameters are associated with positive outcomes. The criteria for specific points used could not be set because these varied between textbooks.

When these hypotheses were tested, no association was observed between the number of points used and positive outcome, but a statistically significant association ($P<0.05$) was found between the total number of treatments given and a positive outcome, even when controlling for methodological quality of the trials. Virtually no trial that administered less than six acupuncture treatments achieved a positive outcome. Although these significant findings suggest association and not causation, the findings may be a starting point to examine dose-response relationships in pilot tests prior to conducting larger trials.

There is no consensus about how to assess treatment adequacy. In the Cochrane protocol for manual lymphatic drainage (MLD)[39] it is suggested that two MLD-certified therapists blinded to study results assign a holistic score of 'adequate', 'inadequate' or 'not enough information to decide' to each of the treatment regimens based on their clinical experience. This simple method has been used in the Cochrane review of acupuncture point stimulation for chemotherapy-induced nausea and vomiting[40] and has been found to attract high inter-rater agreement.

Initially, the Cochrane peer review disagreed with this approach, suggesting the use of explicit criteria for each treatment dimension such as number of treatments, duration of treatment session and frequency of treatments. It was explained that this method had been ruled out given Linde's

lack of enthusiasm after having tried it. Ultimately, the Cochrane peer review accepted the holistic scoring method.

Under-reporting is a common barrier to assessing treatment adequacy. Indeed, data on *de qi* are reported so infrequently in acupuncture trials that it cannot be meaningfully assessed.[38] Similarly, so few treatment details are reported in massage trials[41] that rather than assessing the treatments it is perhaps beneficial to report and discuss the under-reporting problem with hopes of influencing reporting practices in the future.

Treatment adequacy is particularly difficult to assess in mind-body therapies such as meditation. Although meditation is a self-administered therapy, unlike other self-administered therapies such as ingestible substances, compliance with meditation practice does not ensure treatment adequacy. Since compliance is not a good proxy for treatment adequacy, researchers have sought physiological measures that can serve as proxies. However, changes in heart rate or respiratory rate can be achieved by a variety of activities, even reading silently, and, therefore, cannot be used to assess treatment adequacy in meditation trials.[42] Treatment adequacy in CAM ingestible substances requires standardised samples containing active ingredient(s). The predominant limitation of existing botanical trials cited in systematic reviews is the lack of standardised samples.

Practitioner qualifications

Evaluating efficacy means testing an intervention under optimal conditions. This does not guarantee that the benefits of the intervention will carry over in the real world uses (effectiveness), but the efficacy principal does give an intervention the best opportunity to prove itself. Optimal conditions require not only adequate doses, but also optimally qualified and skilled practitioners, in practitioner-based modalities such as massage, acupuncture and chiropractic. Anyone who has visited more than one practitioner of the same modality knows that practitioners' skills vary widely. If an intervention is administered by a less than highly skilled practitioner the trial may be assessing effectiveness rather than efficacy.

Commentators have highlighted three dimensions that relate to practitioner quality: credentials, experience and hands-on proficiency.[43] Ter Riet et al.[44] felt that the practitioner qualifications played such an important role in chronic-pain outcomes for acupuncture that they attempted to use practitioner credentials and practitioner experience as a proxy for adequacy of treatment. However, these details were so seldom reported that the reviewers could draw no conclusions. Haraldsson et al.[41] noted the same under-reporting of practitioners' qualifications in a massage review.

The under-reporting of practitioner qualifications may reflect the lack of serious consideration given to this issue in trial planning. While this issue extends to all practitioner-based modalities, the issue is well illustrated by examples in massage research. Against a backdrop of trials that give no details of practitioner selection, some trials stand out for their conscientious consideration of practitioner selection. The multi-centre 'Relieving End-of life Symptoms with Massage' (REST) study (work in progress at the time of writing) has hired only certified massage therapists with prior experience treating dying patients. Similarly, in a 'massage for low back pain' trial, Cherkin et al.[45] required not only certification and prior experience treating low back pain, but also a 'hands-on working interview' in which researchers received a massage from each therapist prior to being hired.

In systematic reviews, the problem arises as to how massage trials which have made efforts to use optimally qualified and skilled practitioners can be compared with trials where massage has been administered by massage students, chiropractic assistants with no formal massage training or nurses trained in massage only for the study but with no prior experience. Can these latter trials really be considered efficacy trials? At the simplest level, one might do a subgroup analysis

comparing trials of more versus less qualified practitioners. However, this issue is far from resolved. Some suggest that the need for optimally qualified and skilled practitioners varies from condition to condition. For example, some believe that nurses with no formal massage training other than for the purposes of the study are sufficient to administer very basic, formulaic massage to premature infants.[46] Presently, the vast majority of reviews are left to comment on the lack of reporting of practitioner selection criteria in the hope of both raising standards for reporting and, most importantly, raising standards of selection during planning of trials.

Cointerventions

CAM, by definition, is often an add-on treatment. Yet, it is often impossible in CAM systematic reviews to assess cointerventions because they have not been adequately documented in the clinical trials. Nevertheless, it remains that when a conventional treatment is universally and simultaneously used with CAM by a study population and that intervention becomes more effective over time, it can alter the relative benefit of the CAM treatment. In such cases, the impact of the cointervention must be considered.

For example, the 1998 National Institutes of Health (NIH) Consensus statement[47] concluded that acupuncture was effective for chemotherapy-induced illness. However, the most recent (and most efficacious) generation of antiemetics (5-HTJ inhibitors such as ondansetron and granisetron) was only just beginning to be widely used at the time of this statement. When the review of acupuncture for chemotherapy-induced nausea and vomiting began shortly after the NIH conference, there was a question as to whether these more highly efficacious antiemetics would change the relative contribution of acupuncture. Thus, the review needed to take into account the patients' use of antiemetics while receiving acupuncture treatment. Antiemetic regimens are determined based on the emetogenicity rating of the chemotherapy. A rating is achieved through an oncologist assessing the chemotherapy, the emetogenicity ratings and the compliance with modern antiemetic guidelines. Trial investigators provided missing data and subgroup analysis was performed comparing modern and older antiemetics. The results showed that all acupuncture trials gave concomitant antiemetics, and the pooled acupuncture results showed a protective effect. However, no acupuncture trial had given antiemetics wholly consistent with current standards. The positive findings were a 'proof of principle' of acupuncture's effectiveness but could not answer whether acupuncture added benefit to modem antiemetics.[40]

Interpreting results

Generally by the time the review's conclusions are written, reviewers agree on data interpretation. However, such agreement is not always the case. In the 'acupuncture for chemotherapy-induced nausea and vomiting' review, one methodologist interpreted the results differently from the other participating methodologists. While one reviewer suggested that acupuncture may be beneficial for refractory patients, others argued no trials had explicitly assessed refractory patients. The point of the dissenting methodologist was that refractory patients were out of options, that acupuncture is safe and, based on the proof of principle from high-quality randomised trials, it may be beneficial. Reviewers spent an additional month consulting external experts - oncologists, oncology biostatisticians and epidemiologists - as well as discussing the issue internally. After much discussion, the decision was made to not suggest acupuncture for refractory patients but to suggest further trials need to be conducted.[40]

Safety

Safety, not just efficacy, needs to be addressed systematically. Safety data on CAM dietary supplements are particularly challenging due to their lack of government regulation in certain

countries where there is high use such as the USA. As such, adverse-events documentation relies on self-initiated reporting, resulting in under-reporting. Incidence rates of adverse events, therefore, cannot be directly calculated for ingestible substances because the numerator (number of adverse events) is under-reported, and the denominator (number of persons using the substance) is extremely difficult if not impossible to estimate.

For safety to be evaluated in systematic reviews, assessments need to go beyond the clinical trials, which are notoriously small, and into other sources such as drug-interaction databases, population-based surveys and manufacturers' records. Cost, however, remains an obstacle. Some reviews have an explicit and transparent method for assessing adverse effects[48] although most lack this level of detail. Increasingly, CAM reviews at least cite contraindications and possible side effects even when they lack a search of adverse events.

Incidence rates of adverse events of practitioner-based modalities are easier to approximate because they can be assessed through large prospective studies of clinical practices with the consecutive patients providing the denominator. This method has been used in acupuncture,[49,50] massage,[51] and chiropractic.[52,53] Systematic reviews of the prospective studies have also been done.[49]

SUMMARY

Systematic reviews are an evolving science and need to have the same methodological rigor as any other study design. In addition to the methodological issues that apply to all reviews, CAM systematic reviews often have to address additional methodological issues given the complexity of CAM interventions. CAM systematic reviews are particularly valuable not only for their summaries of the evidence but also for the way in which they provide valuable information to guide subsequent clinical research

REFERENCES

1. Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA* 2005;293(19):2362-6.
2. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490.
3. Manheimer E, Berman B. The Cochrane Column: Exploring, evaluating, and applying the results of systematic reviews of CAM therapies. *Explore: The Journal of Science and Healing* 2005;1(3):210-14.
4. Lehmann HP, Goodman SN. Specifications for formalizing clinical significance. *Med Decision Making* 1995;15:424.
5. Dickersin K, Manheimer E. The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin Obstet Gynecol* 1998;41(2):315-31.
6. Bero LA, Jadad AR. How consumers and policymakers can use systematic reviews for decision making. *Ann Intern Med* 1997;127(1):37-42.
7. Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. *Ann Intern Med* 2004;141(12):901-10.
8. Ezzo J, Hadhazy V, Birch S, Lao L, Kaplan G, Hochberg M, Berman B. Acupuncture for osteoarthritis of the knee: a systematic review. *Arthritis Rheum* 2001;44(4):819-25.
9. McNeely ML, Magee DJ, Lees AW, Bagnall KM, Haykowsky M, Hanson J. The addition of manual lymph drainage to compression therapy for breast cancer related lymphedema: a randomized controlled trial. *Breast Cancer Res Treat* 2004;86(2):95-106.
10. Tulder MWV, Cherkin DC, Berman B, Lao L, Koes BW. Acupuncture for low back pain. *Cochrane Database Syst Rev* 1999(2):CD001351.
11. Furlan A, Tulder M, Cherkin D, Tsukayama H, Lao L, Koes B, Berman B. Acupuncture and dry-needling for low back pain. *Cochrane Database Syst Rev* 2005(1):CD001351.
12. Manheimer E, White A, Berman B, Forys K, Ernst E. Meta-analysis: acupuncture for low back pain. *Ann Intern Med* 2005;142(8):651-63. Erratum in: *Ann Intern Med* 2005 Jun 7;142(11):950-1.
13. Diamond BJ, Johnson SK, Torsney K, Morodan J, Prokop BJ, Davidek D, Kramer P. Complementary and alternative medicines in the treatment of dementia: an evidence-based review. *Drugs Aging* 2003;20(13):981-98.
14. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337(8):536-42.
15. Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J R Soc Med* 2003;96(1):17-22.
16. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;309(6964):1286-91.
17. Pilkington K, Richardson J. Exploring the evidence: the challenges of searching for research on acupuncture. *J Altern Complement Med* 2004;10(3):587-90.
18. Murphy LS, Reinsch S, Najm WI, Dickerson VM, Seffinger MA, Adams A, Mishra SI. Searching biomedical databases on complementary medicine: the use of controlled vocabulary among authors, indexers and investigators. *BMC Complement Altern Med* 2003;3(1):3.
19. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276(8):637-9.
20. Savoie I, Helmer D, Green CJ, Kazanjian A. Beyond Medline: reducing bias through extended systematic review search. *Int J Technol Assess Health Care* 2003;19(1):168-78.
21. Avenell A, Handoll HH, Grant AM. Lessons for search strategies from a systematic review, in The Cochrane Library, of nutritional supplementation trials in patients after hip fracture. *Am J Clin Nutr* 2001;73(3):505-10.
22. Pilkington K, Boshnakova A, Clarke M, Richardson J. "No language restrictions" in database searches: what does this really mean? *J Altern Complement Med* 2005;11(1):205-7.
23. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7(1):1-76.
24. Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;31(1):115-23.
25. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000;53(9):964-72.
26. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technol Assess* 2003;7(41):1-90.
27. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. *J Clin Epidemiol* 2005;58(8):769-76.
28. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323(7303):42-6.

29. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005;330(7499):1053.
30. Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. *J Clin Epidemiol* 2006;59(2):107-13.
31. MacPherson H, White A, Cummings M, Jobst KA, Rose K, Niemtzow RC. Standards for Reporting Interventions in Controlled Trials of Acupuncture: the STRICTA recommendations. *J Altern Complement Med* 2002;8(1):85-9.
32. Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 2005;330(7501):1202-5.
33. Deeks JJ, Higgins, JPT, Altman DG, editors. Analysing and presenting results. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.4 [updated March 2005]; Section 8.
34. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine* 2003;28(12):1290-9.
35. Linde K, Jobst K, Panton J. Acupuncture for chronic asthma. *Cochrane Database Syst Rev* 1996(2):CD000008.
36. Molsberger A, Bowing G. [Acupuncture for pain in locomotive disorders. Critical analysis of clinical studies with respect to the quality of acupuncture in particular]. *Schmerz* 1997;11(1):24-9.
37. Birch S. *An exploration with proposed solutions of the problems and issues in conducting clinical research in acupuncture*. Exeter: University of Exeter, 1997.
38. Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;86(3):217-25.
39. Howell D, Ezzo J, Tuppo K, Bily L, Johannson K. Complete decongestive therapy for lymphedema following breast cancer treatment. (Protocol). *Cochrane Database Syst Rev* 2002(1).
40. Ezzo JM, Richardson MA, Vickers A, Allen C, Dibble SL, Issell BF, Lao L, Pearl M, Ramirez G, Roscoe J, et al. Acupuncture-point stimulation for chemotherapy-induced nausea or vomiting. *Cochrane Database Syst Rev* 2006(2):CD002285.
41. Haraldsson BG, Gross AR, Goldsmith CH, Myers CD, Ezzo JM, Morien A, Peloso P. Cervical Overview Group. Massage for mechanical neck disorders. (Protocol). *Cochrane Database Syst Rev* 2004(3).
42. Caspi O, Burleson KO. Methodological challenges in meditation research. *Advances in mind-body medicine* 2003;21(1):4-11.
43. Eisenberg DM, Cohen MH, Hrbek A, Grayzel J, Van Rompay MI, Cooper RA. Credentialing complementary and alternative medical providers. *Ann Intern Med* 2002;137(12):965-73.
44. ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990;43(11):1191-9.
45. Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, Deyo RA. Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. *Arch Intern Med* 2001;161(8):1081-8.
46. Field T, Scafidi F, Schanberg S. Massage of preterm newborns to improve growth and development. *Paediatric Nursing* 1987;13:385-7.
47. NIH Consensus Conference. Acupuncture. *JAMA* 1998;280(17):1518-24.
48. Mulrow C, Lawrence V, Jacobs B. Milk thistle: effects on liver disease and cirrhosis and clinical adverse effects. Evidence Report/Technology Assessment No. 21, AHRQ Publication No. 01-E025. Rockville, MD: Agency for Healthcare Research and Quality. October 2000.
49. Ernst E, White AR. Prospective studies of the safety of acupuncture: a systematic review. *Am J Med* 2001;110(6):481-5.
50. MacPherson H, Thomas K, Walters S, Fitter M. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;323(7311):486-7.
51. Cassileth BR, Vickers AJ. Massage therapy for symptom control: outcome study at a major cancer center. *J Pain Symptom Manage* 2004;28(3):244-9.
52. Cagnie B, Vinck E, Beernaert A, Cambier D. How common are side effects of spinal manipulation and can these side effects be predicted? *Man Ther* 2004;9(3):151-6.
53. Haas M, Goldberg B, Aickin M, Ganger B, Attwood M. A practice-based study of patients with acute and chronic low back pain attending primary care and chiropractic physicians: two-week to 48-month follow-up. *J Manipulative Physiol Ther* 2004;27(3):160-9.

Chapter 6. Published reports of acupuncture trials showed important limitations

Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. J Clin Epidemiol 2006 Feb;59(2):107-13.

ABSTRACT

OBJECTIVE: Systematic reviewers generally evaluate RCTs based on the published reports. We evaluated whether the description of methods in the published reports is an accurate and complete reflection of study procedures used.

STUDY DESIGN AND SETTING: The authors of 51 RCTs included in a systematic review of acupuncture for chronic pain were sent a brief survey that included questions related to the following three important study quality dimensions: 1) generation of allocation sequence, 2) allocation concealment, and 3) blinding of outcomes assessor.

RESULTS: We received 35/51 responses for an overall response rate of 68.6%. Of 35 studies described as randomized in published reports, associated survey responses indicated that four actually used quasi-randomized methods. Among published reports with missing information on these quality dimensions, 27/32 studies used adequate methods for the generation of allocation sequence, 13/34 used adequate allocation concealment and 2/10 were blinded, according to survey responses. Survey responses generally confirmed information about randomization and blinding already described in investigators' RCT publications.

CONCLUSION: Surveying RCT investigators uncovered some information about study quality dimensions not described in published reports.

BACKGROUND

Randomized controlled trials are the gold standard study design for evaluating the effects of healthcare therapies. Because RCTs generally estimate treatment effects with less bias than do observational studies[1-4] many or most systematic reviews limit inclusion to RCTs only. However, the quality of RCTs is not uniform. Several studies have now documented that RCTs of lower quality, as reflected in specific flaws in study design or execution, result in larger, and presumably inflated, estimates of treatment effects compared with RCTs of higher quality.[5-7] Systematic reviewers therefore often evaluate RCT quality and then incorporate quality in the interpretation of results, using sensitivity analysis, subgroup analysis, or an exploration of heterogeneity. In addition, sometimes lower quality studies are excluded from a review.[8-13]

The most commonly used components for assessing quality are, in decreasing order, allocation concealment, any type of blinding, loss to follow-up, intention to treat analysis, and generation of allocation sequence.[12] Various quality scales are used that incorporate these and other quality components.[11,14] Whether quality is evaluated based on the individual quality components or quality scales, such evaluations are generally made based on information from the published report. A major disadvantage of assessing quality based solely on the published report is that this relies on a critical and unexamined assumption that what is written in the report reflects actual procedure and that "if it is not reported, it probably was not done." This study aims to evaluate whether RCT reports provide an accurate and complete description of the procedure and conduct of the RCT by undertaking a survey to determine whether evaluation of key quality dimensions based on the RCT report would change if this report data were supplemented with additional information supplied directly through contact with investigators.

METHODS

Our sample of reports encompassed 51 RCTs that were included in a systematic review of acupuncture for chronic pain.[15] This review was selected because it includes a large number of trials (especially for the field of complementary and alternative medicine) and also because it was the review under investigation at the time the survey was undertaken.

The investigators associated with the 51 RCTs were surveyed to find out whether the following three important quality dimensions not described in their published reports were included in their studies: 1) method of random allocation, 2) allocation concealment, and 3) blinding of outcomes assessor (see Questionnaire appendix). Investigators who already described these quality dimensions in their published reports were also surveyed to determine whether information described in published reports would be corroborated by investigators' survey responses.

All of the published RCT reports included in our systematic review were required to use a variant of the term 'random' to describe the method of generation of allocation sequence; however, information about the appropriateness of this method of allocation sequence generation was not usually included in the published RCTs. The first question in our survey therefore assessed the adequacy of the generation of allocation sequence. The response choices "table of random numbers", "a computerized program", or "drawing numbers from a hat" were considered adequate. In contrast, if the respondent indicated that "alternate assignment" was used, the study was considered not randomized.

The second survey question related to adequacy of allocation concealment. Allocation was considered concealed if the respondent indicated that the method of assigning participants to treatments was to "call a central office when a person was deemed eligible", or to use "sealed envelopes sequentially numbered with assignment inside". In contrast, if the respondent indicated that a "master list generated ahead of time, kept in house and referred to when someone entered the study" was used, the randomization process was not considered concealed.

The final question on our survey related to outcomes assessor blinding. We limited the evaluation of outcomes assessor blinding to trials that did not rely exclusively on patient self-reports, as indicated by the responses to survey question 3 (see Appendix). When only patient self-reports are used, outcomes assessor blinding is not possible: the patients know whether or not they received acupuncture. The only exception to this is acupuncture trials that use a sham acupuncture control group, which is almost universally applied for the primary purpose of blinding the patient to the treatment received.

The RCT investigators were first contacted by mail in March 1998 to learn whether details related to randomization and blinding not described in published reports were part of trial procedure. The response rate to the 1998 mailing was only 49%. To increase our response rates, in April/May 2005 we re-contacted, primarily by e-mail, non-responders to the 1998 mailing. Current e-mail addresses were abstracted from these investigators' recent publications, identified through PubMed and electronic journal databases. These 2005 e-mail contacts were followed up by at least two additional reminder e-mails, faxes or telephone calls. If first authors did not respond, second or final authors were contacted.

Our original objective for the 1998 mailing was to learn about underreporting of randomization and blinding in trials. Based on a March 2005 peer reviewer's comment, we modified the objective for the April/May 2005 follow-up mailing to assess for both under- and over-reporting. To assess for over-reporting, for the 2005 mailing we re-contacted nine investigators who had already responded to the 1998 mailing, but whose 1998 questionnaires did not include the question about outcomes assessor blinding because this quality dimension was already reported in their associated publications. In addition, for the 2005 mailing, we sent surveys to eight investigators not included in the 1998 mailing, either because their associated publications already adequately described both randomization and outcomes assessor blinding, or because their publications had not yet been identified for the systematic review in 1998.

This survey was originally conducted in 1998 during the preparation of a systematic review of acupuncture for chronic pain.[15] For this systematic review,[15] a best evidence synthesis was used to formulate conclusions on the effectiveness of acupuncture for chronic pain, as compared with four different kinds of controls (i.e., waiting list, inert control, sham acupuncture, active control). A best evidence synthesis takes into account study outcomes as well as study quality in grouping conclusions about efficacy into "levels of evidence". Study quality was assessed in this review by scoring the published reports according to the validated Jadad scale.[16] We calculated whether changes in the Jadad overall quality scores resulting from contacting authors would produce a change in the level of evidence of acupuncture for chronic pain, as compared with any of the control groups. This re-analysis involved the rescoring of only a single criterion -- generation of allocation sequence – which was in common between our survey questions and the validated Jadad scale. We also evaluated whether excluding the trials found to use quasi-randomized methods, according to the results of the survey, would affect the results of the systematic review. Finally the systematic review[15] included sensitivity analyses to test whether individual methodological quality components were significantly ($p<.05$) associated with positive study outcomes in the published RCTs. Supplementing the published RCT quality data with investigator survey responses, we reanalyzed the data to determine whether any statistically significant associations found between any of the three quality dimensions surveyed on and positive outcomes were affected by the new information obtained by the survey responses.

RESULTS

We received 35/51 responses for an overall response rate of 68.6%. All but three responses were received from first authors. Despite extensive searches for valid investigator contact information

using PubMed, electronic publication databases, and the Internet, no contact information could be obtained for six investigators. Ten other investigators were successfully contacted but they did not respond to our survey. Difficulties in locating investigators and obtaining responses could potentially be explained by the fact that a large proportion of the RCTs included in our sample were conducted in earlier decades (8 studies from the 1970s; 24 studies from the 1980s; 19 studies from the 1990s).

Of 35 studies described as randomized in published reports, associated survey responses indicated that four actually used quasi-randomized methods. Among published reports with missing information on these quality dimensions, 27/32 studies used adequate methods for the generation of allocation sequence, 13/34 used adequate allocation concealment and 2/10 were blinded, according to survey responses (see Tables 1-3). The most commonly used method for generation of the random allocation sequence was a table of random numbers, while the most commonly used method for achieving allocation concealment was the use of sealed sequentially numbered envelopes (see Table 4).

Among investigators who already described these quality dimensions in their associated publications, 3/3 confirmed in their survey responses adequacy of generation of allocation sequence, 1/1 confirmed adequacy of allocation concealment and 17/18 confirmed outcomes assessor blinding (see Tables 1-3). The single investigator who reported outcomes assessor blinding in his publication but not in his survey response indicated in his survey response fax cover sheet that "As you appreciate, this is a long time ago and my memory is not what it was! I have completed your questionnaire as far as I recall and it follows".

Recalculating the Jadad scale overall quality scores based on the published reports supplemented with the survey responses affected the level of evidence of acupuncture, as compared with the waiting list control group only. The evidence for acupuncture's superiority to a waiting list control was changed from 'limited' to 'strong' as a result of the investigator contacts. The results of the systematic review were not affected by excluding the four quasi-randomized studies. In addition, the statistical significance of the association between positive results and any of the three quality dimensions was not affected by the survey results.

Table 1. Comparison of generation of allocation sequence in RCT publications and corresponding investigator survey responses

Generation of allocation sequence	RCT publication	
	Adequate	Unstated*
Investigator survey response		
Adequate	3	27
Inadequate	0	4
Unknown	0	1

*All of the RCT publications in our sample were required to use a variant of the term 'random' to describe the method of generation of allocation sequence; however, information about the adequacy of the method of allocation was usually not stated in the RCT publications.

Table 2. Adequacy of allocation concealment in RCT publications and corresponding investigator survey responses

Allocation concealment	RCT publication	
	Adequate	Unstated
Investigator survey response		
Adequate	1	13
Inadequate	0	20
Unknown	0	1

Table 3. Adequacy of outcomes assessor blinding in RCT publications and corresponding investigator survey responses

Outcomes assessor blinding	RCT publication	
	Adequate*	Unstated
Investigator survey response†		
Adequate	17	2
Inadequate	1	8
Unknown	0	0

*Two investigators who had already responded to the 1998 questionnaire, but whose 1998 questionnaires did not include the question about outcomes assessor blinding because this quality dimension was already reported in their published studies, did not respond to the 2005 mailing on outcomes assessor blinding.

†Five investigators indicated that outcomes assessment was exclusively by self-report, and were not included in this analysis.

Table 4. Methods used for the generation of allocation sequence and allocation concealment among published RCTs with missing information on these quality dimensions*

Generation of allocation sequence (n=32)	
Table of random numbers	17
A computerized program	3
Drawing numbers from a hat	7
Alternate assignment	4
Other	
Allocation concealment (n=34)	
Called a central office when person deemed eligible	1
Sealed sequentially numbered envelopes	12
Master list kept in house	12
Other†	8

*One respondent could not remember the method of generation of allocation sequence or the method of allocation concealment.

†For the 'Other' response choice category, four respondents wrote in that alternate assignment was used and four respondents wrote in that unnumbered, shuffled and selected envelopes were used.

DISCUSSION

Information critical to the evaluation of quality is often missing from RCT reports. This omission presents a dilemma for systematic reviewers who must decide whether to contact the trial investigators to obtain the data related to quality or whether instead to assume that the methodological details related to quality not described in the report were most likely not part of the study's procedure. Generally, the latter option is chosen, both because of the time and effort

required to contact investigators and because of the lack of evidence to show that this effort spent would substantially impact the results of a review.

Our survey results suggest that trial publications not specifying the method of generation of allocation sequence used quasi-randomized methods in 12.5% of cases. Among published reports with missing information on allocation concealment and outcomes assessor blinding, 13/34 studies used adequate allocation concealment and 2/10 studies were blinded, according to survey responses.

Many studies have suggested a lack of rigor in RCT conduct based on inadequate reporting of methodological detail in published articles. For example, Moher et al found in 1998[6] that only 15% of a sample of RCT reports clearly described the generation of the allocation sequence, and only a fraction of these reported that allocation concealment was used. Among the articles in our sample, only 3/35 (8.6%) described the generation of allocation sequence, and only one of these reported that allocation concealment was used. Our study results suggest that when an RCT publication omits a description of allocation concealment and blinding, these procedures are often not part of the conduct of the RCT.

The reports included in our sample were published between 1974 and 1997, largely before the introduction of the CONSORT statement in 1996.[17] This was also before STRICTA guidelines, an adaptation of CONSORT specifically for acupuncture trials. Before 1996, investigators who used adequate methods for randomization and blinding may not have realized the importance of fully describing these methods in the report,[18] and editors may not have realized the importance of publishing this information. With the widespread dissemination of CONSORT through hundreds of journals, more authors and editors are now aware of RCT reporting guidelines[19]; thus, we would suspect that for trials published today, an even greater probability exists that an unreported procedure was not carried out at all.

A possible limitation of our study relates to the lack of specificity of the possible response choices provided to the allocation concealment question. Two respondents explained that their assignment of participants to treatment groups was similar to that specified in the response "sealed envelopes sequentially numbered with assignment inside," but noted that the envelopes they used, while shuffled and selected, were not numbered. Based on our correspondence with an expert on modern research on trial quality [personal communication Kenneth Schulz, 21 July 2004], we decided that these responses should not be classified as concealed allocation. Also, the 'sealed envelopes' response choice did not specify that the envelopes must be opaque; some researchers[7] would consider opaqueness of the envelopes to be an additional requirement for truly concealed allocation, while other researchers[5] would not. Another response choice for this question -- "master list generated ahead of time, kept in house and referred to when someone entered the study" -- was not considered concealed allocation. However, if such a master list were maintained with the biostatistician, then this method would be concealed allocation. Yet we did not allow for this possibility with our response choice wording. While we used pre-defined response choices together with an additional open-ended response choice, if we had instead used exclusively open-ended questions we may have obtained more precise information about the methods used in each of the trials. However, using only open-ended questions would have had the disadvantage of difficulties in interpretation and coding of responses and also possibly a lower response rate, as a result of the additional time and effort required for the respondents to complete the surveys.

Another possible limitation relates to the fact that we did not obtain a response from 16 investigators. Selection bias may have affected our findings if these 16 non-respondents were more or less likely than the 35 respondents to have used appropriate methods for randomization and blinding in their trials but failed to document these methods in their RCT publications.

While our survey was brief, we believe that it captured the most important quality elements of an RCT, and that the collection of additional information would have resulted in a more complicated survey, as well as—perhaps--a lower response rate. We collected information on three of the most commonly used criteria for assessing RCT quality[12,20]: allocation concealment, blinding, and generation of allocation sequence. Allocation concealment is suspected to be the study design factor most strongly associated with internal validity.[3,5-7,10] Blinding has also been associated with internal validity in empirical studies.[5,7] In this survey, we chose to collect information only on outcomes assessor blinding despite the fact that caregiver blinding and patient blinding are also included as items in many criteria lists for RCT quality assessments.[21] While different investigators have different interpretations of who should be blinded,[22] we surveyed only on outcomes assessor blinding because this is a particularly important measure for avoiding bias when evaluating a subjective outcome measure, such as pain, which was the outcome used for the RCTs in our sample.[7,23-25] We did not collect information on caregiver blinding because it is virtually impossible to blind acupuncturists on whether or not they are administering true acupuncture to their patients. We also did not ask about patient blinding because this is generally only possible in acupuncture trials that use a sham acupuncture control group, which is used for the primary purpose of blinding patients to the treatment received. Although the evidence of an association between generation of allocation sequence and study validity is less clear and not consistent across studies,[5-7,26] we also included a question about this in our survey because allocation concealment depends in large part on an adequate generation of allocation sequence.[7,10]

While loss to follow-up and intention to treat analysis are other commonly used components for assessing quality in published RCTs,[12] we did not survey on either of these items. Loss to follow-up was not surveyed on because no consistent relation has been shown between what is reported on loss to follow-up and bias in estimating treatment effects.[5,7] In addition, survey respondents could not be expected to answer a general yes or no question reliably and impartially about the contribution of dropouts towards biasing their RCT results, particularly considering that nearly all studies have some participants drop out, and the magnitude of attrition, the differences in attrition between study groups, and the times of attrition across the groups may all contribute to the effects of attrition on study validity. It is unlikely that respondents would have remembered details about attrition, considering that over half of the RCT reports in our sample were published more than ten years prior to the date of the survey. Even for published reports, the Cochrane Handbook suggests that reviewers should be "cautious about using reported follow-up as a validity criterion, particularly when it is implied rather than explicitly reported".[27] We also did not survey on the intention to treat criteria because there is confusion among researchers about its definition and lack of agreement about its correct application.[28]

For two of the studies in our sample of respondents, we abstracted data from published abstracts, because full reports had not yet been published for these studies by 1998, when the data abstraction was completed. The full reports of these studies, published later, in 1999 and 2001, included details, missing from the abstracts, on generation of allocation sequence, allocation concealment, and (for one of the two) outcomes assessor blinding. The inclusion of these two abstracts in our sample may have resulted in our having slightly overestimated the amount of missing methodological quality data obtained by surveying investigators. For example, if we had abstracted data from the full reports associated with these two studies, instead of from the abstracts, we would have found that, among studies with missing information on these items, 25/30 studies used an adequate generation of allocation sequence, 11/32 used allocation concealment, and 1/9 used outcomes assessor blinding.

Two other studies[29,30] have investigated whether changes in RCT quality assessments resulted from contacting investigators. Liberati et al abstracted information on multiple design and report

features for each RCT in their sample of trials of primary treatment for breast cancer. They created unvalidated scaled scores for each study using this data abstracted, and then assessed whether additional information obtained directly from the investigators would have changed these scores. Their study found that information obtained directly from the investigators added a mean of 7% to their validity scores.[30] Because the fifteen items comprising their scale were merged into a single total score, it is impossible to tell whether items empirically shown to be related to validity (e.g., blinding, randomization) were responsible for the small internal validity score increases or whether the score increases were instead due to items related more to other design features, such as, for example, precision (e.g., sample size validity item) rather than to validity. Regardless, the score increases were small, leading the authors to conclude that poor reporting does reflect faulty methods.

Hill et al[29] surveyed investigators of rheumatology trials published in 1997-1998. They found that among articles with unreported methods for random sequence generation or allocation concealment, almost 80% used adequate methods for each. Their results may have differed from ours because of the different topic areas covered (i.e., acupuncture vs rheumatology specific) or the different publication year periods of the reports sampled. The sample covered by Hill et al was mostly drug trials. For the subset of non-drug trials in their sample, survey results showed that inappropriate methods were used much more frequently. One finding in common between our two studies is that Hill et al also found that a relatively large proportion of reports (7/29) that were described as randomized were not truly randomized.

Soares et al[18] compared published trial reports with their associated protocols in order to determine the likelihood that important study procedures were included in the protocol (and consequently the trial) but not in the published report. They found that the study procedures (e.g., allocation concealment) that had actually been done very often went unreported in the published articles. However, Soares's sample did not include a mix of studies from independent trialists, but rather restricted to trials of the Radiation Therapy Oncology Group, a cooperative oncology group sponsored by the US National Cancer Institute. Protocols published by this group must undergo a rigorous peer review and approval both by its own committee system and by the National Cancer Institute, thus increasing the likelihood that the protocols will report necessary design features. The majority of the reports included in this study's sample were published before the introduction of CONSORT and thus prior to the current standard required by most journals of detailed reporting of design features in trial reports. Thus, while these trialists would have likely included relevant study design information for their peer-reviewed protocols, they may not have reported these methodological details in the final publication. In addition, the Radiation Therapy Oncology Group had conducted early research on quality dimensions of randomized trials,[31] and therefore their knowledge of important procedures to implement in trials may not have been representative of all trialists. A recent study using a more representative sampling of trialists has shown that most RCT reports with unclear allocation concealment also had inadequate allocation concealment according to the study protocol, thus suggesting that inadequate reporting did in fact reflect inadequate methods.[32] McGrath et al[33] reported that contacting investigators resulted in obtaining at least some new data for 17% of the 133 RCTs relevant to the meta-analyses they were conducting; however, they did not study whether this additional data affected their meta-analysis results.

Some systematic review methodologists have advocated contacting investigators to obtain information missing from RCT reports.[34] While our survey suggests that contacting trialists may result in obtaining previously unpublished information about methodological quality, the potential gains of obtaining the missing information may be outweighed by the reporting bias such efforts may introduce, as data obtained directly from investigators has not been peer-reviewed and thus may not be as reliable as data extracted from published articles.

CONCLUSION

The international Cochrane Collaboration has already prepared 2,356 systematic reviews (as of Issue 2, 2005), and another 1,569 are in the protocol stage. To efficiently maintain this prodigious rate of preparing the reviews that serve as the backbone of evidence-based medicine, it is important to investigate which aspects of review preparation are required for the improvement of review quality. Existing studies covering different topic areas and publication years have had contradictory findings on the utility of contacting investigators to obtain additional information about trial quality. Because making such contacts is time and resource intensive, more research is needed to determine whether it results in the retrieval of new, valid information about study quality.

Appendix. Questionnaire

1) Which method most closely describes how the sequence for treatment group assignment was generated?
__ table of random numbers
__ a computerized program
__ drawing numbers from a hat
__ alternate assignment
__ other (*please describe*):

2) Which method most closely describes how the actual assignment to treatment group was done?
__ called a central office when a person was deemed eligible
__ sealed envelopes sequentially numbered with assignment inside
__ master list generated ahead of time, kept in house and referred to when someone entered the study
__ other (*please describe*):

3) Was there an outcomes assessor in the study, or were all the outcomes self-reports?
__ self reports only
__ outcomes assessor

4) If there was an outcomes assessor in the study, was there a way of blinding the outcomes assessor to the treatment group assignment at the time of the assessment?
__ yes
__ no
__ don't know

REFERENCES

1. Altman DG. Randomisation. *BMJ* 1991;302(6791):1481-2.
2. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? *BMJ* 1999;318(7192):1209.
3. Chalmers TC, Celano P, Sacks HS, Smith H, Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309(22):1358-61.
4. Sacks H, Chalmers TC, Smith H, Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72(2):233-40.
5. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135(11):982-9.
6. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352(9128):609-13.
7. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273(5):408-12.
8. Berlin JA, Rennie D. Measuring the quality of trials: the quality of quality scales. *JAMA* 1999;282(11):1083-5.
9. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995;48(1):167-71.
10. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323(7303):42-6.
11. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care* 1996;12(2):195-208.
12. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005;330(7499):1053.
13. Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol* 2001;54(7):651-4.
14. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16(1):62-73.
15. Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;86(3):217-25.
16. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12.
17. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276(8):637-9.
18. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, Djulbegovic B. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328(7430):22-4.
19. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285(15):1992-5.
20. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280(3):278-80.
21. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine* 2003;28(12):1290-9.
22. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, Bhandari M, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285(15):2000-3.
23. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. *Stat Med* 1989;8(4):441-54.
24. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282(11):1054-60.
25. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. *JAMA* 1975;231(10):1038-42.
26. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287(22):2973-82.
27. Higgins JPT, Altman DG, editors. Assessment of study quality. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 4.2.4 [updated March 2005]; Section 6.
28. Deeks JJ, Higgins, JPT, Altman DG, editors. Analysing and presenting results. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 4.2.4 [updated March 2005]; Section 8.

29. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;55(8):783-6.
30. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986;4(6):942-51.
31. Herson J. Patient registration in a cooperative oncology group. *Control Clin Trials* 1980;1(2):101-10.
32. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005;330(7499):1049.
33. McGrath J, Davies G, Soares K. Writing to authors of systematic reviews elicited further data in 17% of cases. *BMJ* 1998;316(7131):631.
34. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.4 [updated March 2005].

Chapter 7. Meta-analysis: acupuncture for low back pain

Manheimer E, White A, Berman B, Forys K, Ernst E. Meta-analysis: acupuncture for low back pain. Ann Intern Med 2005 Apr 19;142(8):651-63. Erratum in: Ann Intern Med 2005 Jun 7;142(11):950-1

ABSTRACT

BACKGROUND: Low back pain limits activity and is the second most frequent reason for visits to physicians. Previous research shows widespread use of acupuncture for low back pain.

PURPOSE: To assess acupuncture's effectiveness for treating low back pain.

DATA SOURCES: Randomized controlled trials were identified through searches of the following databases through August 2004: MEDLINE, Cochrane Central, EMBASE, AMED, CINAHL, CISCOM, and GERA. Additional data sources included previous reviews and personal contacts with colleagues.

STUDY SELECTION: Randomized controlled trials comparing needle acupuncture with sham acupuncture, other sham treatments, no additional treatment, or another active treatment for patients with low back pain.

DATA EXTRACTION: Data were dually extracted for the outcomes of pain, functional status, overall improvement, return to work, and analgesic consumption; additionally, the quality of the studies was assessed.

DATA SYNTHESIS: The 33 RCTs meeting the inclusion criteria were sub-grouped according to acute or chronic pain, style of acupuncture, and type of control group used. For the primary outcome of short-term relief of chronic pain, the meta-analyses showed that acupuncture is significantly more effective than sham treatment (standardized mean difference -0.54 [95% CI, -0.35 to -0.73]; 7 trials) and no additional treatment (standardized mean difference -0.69 [95% CI, -0.40 to -0.98]; 8 trials). For patients with acute low back pain, data are sparse and inconclusive. Data are also insufficient for drawing conclusions about acupuncture's short-term effectiveness compared with most other therapies.

LIMITATIONS: The quantity and quality of the included trials is variable.

CONCLUSIONS: Acupuncture is effective in providing relief from chronic low back pain. No evidence exists that acupuncture is more effective than other active therapies.

The online version of the thesis Table of Contents includes links to the text of the following related material:

- Supplementary appendices;
- Letter to the Editor, and authors' response:
    - Shekelle P. Meta-analysis: acupuncture for low back pain [Letter to the Editor]. *Ann Intern Med* 2005;143(9):691-2.
    - Manheimer E, White A, Ernst E, Langenberg P. Acupuncture for low back pain [Authors' response]. *Ann Intern Med* 2005;143(9):692.

INTRODUCTION

Low back pain is the most common cause of activity limitation in people younger than 45 years in the United States, and the second most frequent reason for visits to the physician.[1] A 1997 US survey showed that 54% of patients reporting back or neck pain had used a complementary therapy within the past year,[2] while a 2002 study showed back pain to be the most common reason for visits to acupuncturists in the United States.[3] Given this high degree of public interest, the question whether or not acupuncture alleviates low back pain is relevant.

Prior to this study, at least three publications systematically reviewed the primary research up to 1996.[4-6] Their conclusions, limited by the paucity, heterogeneity, and poor quality of the studies, were somewhat discordant. Since then, several new studies have been published. Some of these studies have been reviewed,[7] but the review did not take full account of study quality and did not attempt a meta-analysis of the data.

We decided to undertake a new systematic review and meta-analysis to test the hypotheses that, for the treatment of low back pain, acupuncture is more effective than penetrating and non-penetrating sham acupuncture, than other sham treatments, than active treatments, and than doing nothing. In addition, our review addresses the hypotheses that treatment effect size is correlated with study quality, treatment factors and patient factors.

METHODS

Search

We searched the following computerized databases from their inception until August 2004: MEDLINE, Cochrane Central, EMBASE, AMED, CINAHL, and two specialised European databases CISCOM in London, UK, (to February 2003) and GERA in La Garde, France, which both include 'grey literature' such as unpublished studies and conference reports. Text-word searches were performed of titles and abstracts for acupuncture, electroacupuncture, percutaneous electrical nerve stimulation, AND backache, back pain, low back pain, dorsalgia and lumbago. We contacted experts in Japan who updated their earlier search of Igaku Chuo Zasshi, 1987-2003[8] and also hand-searched the Journal of the Japan Society of Acupuncture and Moxibustion 2000-2004. All studies included in previous reviews[4-7] were reconsidered for inclusion in this review.

In addition, our own files were screened, and experts in the United Kingdom, the United States, Germany, Sweden and Norway were contacted and asked to contribute any studies of which they were aware. Copies of all papers that could be reports of randomized clinical trials (RCTs) or reviews of RCTs were obtained.[9] The bibliographies of all papers thus retrieved were scanned for further references. Unpublished reports were not included.

Study selection

Those articles reporting random assignment of participants to acupuncture or a control group for the treatment of any type of low back pain in human beings were considered for inclusion by AW and EM independently. Studies that included both neck and low back pain patients were excluded from the review unless the data for low back pain patients were available separately. Acupuncture was defined in an inclusive manner to reflect its use in various traditions and theoretical approaches: the intervention had to involve the insertion of needles into the skin, but not for the purpose of injection. This definition includes any intervention that was described by the study author(s) as 'acupuncture', or that was clearly identified as acupuncture in the opinion of the reviewers. Therapies that are similar to acupuncture but do not involve needle insertion (e.g., laser acupuncture, electroacupuncture without needles) were excluded because most authorities believe acupuncture entails needle insertion.[10] Trials in which one form of acupuncture was compared only with another form were excluded. Studies that provided usable data for any of the following

outcome measures were included: pain, measures of functional status (e.g., Oswestry, Roland Disability Questionnaire), overall improvement, return to work, and analgesic consumption. Studies that only reported relief of pain immediately after a single treatment with acupuncture were excluded.

Reports in English, Japanese, Korean, and Chinese, as well as in Germanic (including Scandinavian) and Romance languages, were included; where necessary, translations of essential details were obtained. When more than one publication described a single trial, only one report was included.

Data abstraction

Data were extracted independently by AW and EM onto a piloted spreadsheet, except in the case of non-English language articles, from which data were extracted by an expert and checked in-house. Differences were settled by discussion with reference to the original article. In the few instances where disagreements persisted, an additional independent researcher was consulted who is an expert in review methodology. For the present purposes, three styles of acupuncture and three methods were defined (see Relevant Terms, Table 1).

Table 1. Acupuncture-related terms and their relevance*

Acupuncture is a healing technique that involves the insertion of needles into the body to promote health. It can be traced back at least 2,500 years as part of the healing system in China.

Sham controls in acupuncture research

Sham acupuncture is any intervention designed to make patients believe they are receiving acupuncture. Usually this involves inserting needles superficially and/or at inappropriate sites, and not stimulating them, known as 'penetrating' sham. Occasionally blunt devices are used to apply pressure, without penetration.

Sham transcutaneous electrical nerve stimulation (TENS) involves attaching one end of a TENS electrode to the patient's skin and the other end to an inactivated apparatus. This intervention is designed to at least partially control for the placebo effects of acupuncture treatment.

Styles of acupuncture

Chinese acupuncture: needles inserted into traditional meridian points, usually with the intention of influencing energy flow in the meridian. Additional tender points may also be used.

Japanese acupuncture: superficial needling in the area of the pain and traditional points, using traditional Chinese concepts.

Western acupuncture: the use of unnamed tender or trigger points only, to stimulate nerves or muscles, rejecting traditional concepts of energy and meridians. This treatment may not be accepted as 'acupuncture' by traditional acupuncturists.

Methods of selecting acupuncture points

In **Individual** selection, the practitioner is free to choose any points

In **Formula** selection, the same fixed points are used for all patients

In **Flexible formula** selection, a fixed formula is used and some additional points are chosen according to a patient's tenderness or symptoms

De Qi ("arrival of energy") is a sensation of numbness or distension sometimes generated by stimulating acupuncture needles by hand or with an electrical current. According to acupuncture theory, activation of de qi may be one indication that acupuncture is exerting its beneficial effects.

*We hypothesized that the effects of acupuncture treatment may be correlated with methods of selecting acupuncture points, de qi, and a number of other treatment related factors.

Data were extracted for five outcome categories (pain, functional status, overall improvement, return to full work, analgesic consumption) where available.

Outcomes were extracted for all time intervals reported. Outcomes less than six weeks were considered short-term and the measurement closest to three weeks was used for the meta-analysis. Outcomes more than 6 weeks were considered long-term and the measurement closest to 6 months was included in the meta-analysis. For crossover studies, the risk of carry-over effects was considered prohibitive, so only the first arm of the study was considered.

Quality assessment

The quality of the studies was assessed in two ways. First, using a modification of a validated method[11] a maximum of five points was awarded in three categories: randomization -- two points for an appropriate method, one point if method not described, zero points for an inappropriate method; blinding -- one point for patient blinding, one further point if blinding was tested post-treatment; and, finally, withdrawals and dropouts -- one point if there was a statement either giving full details of withdrawals and dropouts or confirming that there were none. As recommended,[11] a score of 2 points or fewer indicates poor quality.

Secondly, quality was assessed using ten criteria taken from the Cochrane Back Review Group (Table 2),[12] as used in a recent review,[13] extracted by one author (AW) with random confirmation by EM. A score of four points or fewer indicates poor quality.

Table 2. Cochrane Back Review Group Criteria List for Methodological Quality Assessment of RCTs*

Criteria
Was there a randomization method using an adequate procedure?
Was the treatment allocation concealed?
Was the care provider blinded to the intervention?
Was there control for co-interventions?
Were co-interventions reported for each group separately?
Was the patient blinded to the intervention?
Was the outcome assessor blinded to the intervention
Was the withdrawal and dropout rate <20% short term, <30% long term with no substantial bias?
Was the timing of the outcome assessment in both groups comparable?
Did the analysis include an intention-to-treat analysis?

*From van Tulder et al[12]

Quantitative data synthesis

The studies were placed into predefined subgroups according to: 1) acute or chronic low back pain (defined as: acute = three months or less; chronic = more than three months; where not described, a decision was made from the trial setting and recruitment information (e.g., primary care was considered acute; secondary care chronic)); 2) style of acupuncture (see Relevant Terms, Table 1); 3) Control group: sham acupuncture involving needle penetration, non-penetrating sham acupuncture, sham transcutaneous electrical nerve stimulation (TENS) (see Relevant Terms, Table 1), no additional treatment, and other active treatments. The no additional treatment control grouping includes those studies where adjunctive treatment such as physical therapy was given to both the acupuncture and the control group; it also includes those studies where the control intervention consisted of providing patients with educational materials on low back pain.

The primary outcome of the review was quantitative synthesis of the short-term effectiveness of acupuncture on pain in each of the above categories. To test whether sham acupuncture and sham

TENS controlled trials were sufficiently homogeneous to combine for the secondary analyses, we used the I-squared method.[14]

We used the software Comprehensive Meta-analysis [Biostat, Inc. Englewood, NJ] for data analysis. The inverse variance computational model was employed, using the more conservative random effects to account for the expected heterogeneity. Separate analyses were performed for each of the main outcome measures using standardized mean difference (Hedge's g effect size) or odds ratio, depending on the nature of the measure. For the standardized mean difference, a one unit of effect size for pain and for functioning corresponds with a 25 point difference on the visual analogue scale (VAS) and a 6 point difference on the Roland Disability Score. Using standards established by the Cochrane Back Group Editorial Board, a minimum 10-mm difference on the VAS and a 2-point difference on the Roland are considered clinically important.[13] Analyses were performed on short-term endpoint and long-term endpoint, as defined above. Wherever the studies in any of our predefined subgroups were clinically very heterogeneous, or where insufficient data were reported for pooling, we planned to use a narrative synthesis rather than a meta-analysis. We assessed for the likelihood of small study bias using a funnel plot.

In reports where these values were missing, we used the method previously described,[13] in which the median is taken as the mean and the SD is assumed to be 25% of the measure's range. To test the effect of these assumptions, we performed a sensitivity analysis on the short-term pain primary outcome measure, substituting imputed mean values both one SD higher and one SD lower than the imputed values, and then substituting a SD equal to the maximum of any study (80%[15]).

Finally, the I-squared tests were used to evaluate whether there was heterogeneity in the results of the effects of acupuncture at different levels of any of the quality criteria-, treatment-, or patient-related factors. These heterogeneity analyses were conducted separately for the sham controlled and no additional treatment controlled sets of trials on the short-term pain outcome. The cut-points for dichotomising levels of the treatment- or patient-related factors were based on a literature review of earlier clinical trials and systematic reviews, practice patterns of acupuncturists, and Eastern texts. For example, the number of sessions was dichotomized at six or more, versus fewer than six, because an earlier systematic review[16] showed that six or more treatments was associated with better effects of acupuncture. Based on our literature review, we set cut points for treatment factors (individualised vs. formulaic[16]; number of needles[16-18]; obtaining de qi[16,19]; number of treatments[16,17]; length of each session[20]; number of sessions per week; manual vs. electrical stimulation); and patient factors (in primary or secondary/tertiary care; previous surgery or not; leg pain or not).

The funding source had no role in the choice of topic or in the design, conduct, or reporting of the study. They also had no role in the decision to submit the manuscript for publication.

RESULTS

A total of 561 possibly relevant references were considered (see Supplementary Appendix Figure 1), of which full copies were retrieved of 82 papers and five new reviews. Three studies included in previous reviews were excluded: one for insufficient data,[21] one for failing to present separate results for low back pain patients[22] and one for using saline injections rather than acupuncture.[23] In all 33 studies were included in the review.[15,24-55] Ten were from previous reviews,[26-29,34,40,42,43,48,50] sixteen were new studies from our own files or identified through MEDLINE,[24,25,30-33,36,37,39,41,44,45,49,51-53] three were from the GERA database,[15,54,55] one was sent by an acupuncture researcher in Italy,[46] and three were located by the Japanese searches.[35,38,47] Additional details of methodology were obtained from another paper,[56] and unpublished data were obtained from five authors (outcome data[24,33,48] and methodological information[15,46]).

Description of studies

Table 3 describes the main study characteristics and summarizes the findings of the 11 RCTs that could not be combined because of their fundamental heterogeneity. Four RCTs were in patients with acute low back pain with different control interventions, three were in pregnancy-associated low back pain, and four evaluated Western style acupuncture for chronic low back pain, using various control interventions.

Table 4 gives the characteristics of the 22 RCTs of Chinese acupuncture for chronic pain that were included in the meta-analyses. (Supplementary Appendix Tables 1-6 contain further study details.) Eight trials appear to include patients with more severe status, either because of their setting in inpatient care,[40,45] tertiary care[24,46] or rehabilitation unit,[15] or because they included a significant number of patients with leg pain[26,52] or after earlier back surgery.[26,40,41,46] In nine studies, all patients were offered co-interventions, consisting of physical therapy in three,[15,41,45] back exercises in three,[25,36,52] education in one[40] and Chinese massage in two.[54,55] Three studies[29-31] also included a small proportion of patients with higher spinal (but not neck) pain but were nevertheless included in the review. The majority of the studies used well-validated instruments to measure pain[15,24,25,30-32,35,36,40,41,43-47,52] and functional status.[25,30,31,35,41,44,47,52]

Table 3. Study characteristics and results of studies not combined in meta-analysis

Reference, Date	**Quality***	**Acup Type**; no. of sessions; times per week	**N**	**Control** Interventions	**N**	**Any co-interventions** (except drugs) in all groups	**Pain** measure	**Function** measure	**Global score, work or drug intake**	**Measurement points†** (not all outcomes at all points)	**Result** (main comparison)
Acute low back pain											
Duplan et al, 1983[27]	2; 5	Chinese; 5; 5	15	Sham acup	15	-	VAS on standing	-	Analgesic use	eot	Acup sig superior for both outcomes
Kittang et al, 2001[37]	3; 3	Chinese; 4; 2	30	NSAID	30	-	VAS	-	Analgesic use	eot, 3m, 6m	VAS no sig difference; acup sig fewer drugs
Kurosu, 1979[38]	1; 0	Chinese; 3; ?	10	Moxibustion	10	-	Change rating (from -1 to +2)	-	-	eot	No sig difference
Tsukayama et al, 2002[49]	3; 4/5	Chinese; 4; 2	10	TENS	10	-	VAS pain relief	JOA	-	eot	Acup sig superior for pain only

Reference, Date	**Quality***	**Acup Type**; no. of sessions; times per week	**N**	**Control** Interventions	**N**	**Any co-interventions** (except drugs) in all groups	**Pain** measure	**Function** measure	**Global score, work or drug intake**	**Measurement points†** (not all outcomes at all points)	**Result** (main comparison)
Ante-natal low back pain											
Guerriero da Silva, 2004[33]	1; 2	Chinese; 10; 1 or 2	27	Usual care	34	-	Numerical rating scale (average pain)	Numerical rating scale	Global 10 pts (patient); analgesic use; work capacity 10 pts	eot	Acup sig superior for all outcomes
Kvorning, 2004[39]	3; 5	Chinese; 6; 2 then 1	34	Usual care	34	-	Numbers with decrease in pain	Effect of pain on 8 activities (3 pts)	Analgesic use	eot	Acup sig superior for all outcomes
Wedenberg et al, 2000[51]	3; 2	Chinese; 6; 1	30	Physical therapy	30	-	VAS, evening	Disability Rating Index	Global ('good or excellent help')	eot	Acup sig superior for pain and disability
Chronic low back pain, western acupuncture											
Garvey et al, 1989[29]	3; 5	western; 1; n/a	20	a)TP injection b)Spray+ acupressure	a)27 b)16	Showers, activity restriction	-	-	Global 2 pts (patient)	2w	No sig difference between any groups
Gunn et al, 1980[34]	1; 1	western; 8; 2	29	No acup	27	Physical therapy	-	-	Fit for work	eot	Acup sig superior
Macdonald et al, 1983[42]	1; 3	western; 10; 1	8	Sham TENS	9	-	VAS pain, and VAS pain on activities	-	-	eot	Acup sig superior for most outcomes
Yokoyama et al, 2004[53]	2; 2/3	western; 16; 2	20	TENS	20	-	VAS peak pain	Physician assessment (4 pts)	Anti-inflammatory drug use	eot, 1m, 2m	Acup sig superior for all outcomes at eot; no difference at 2m

Abbreviations: acup=acupuncture. eot=end of treatment. JOA=Japanese Orthopaedic Association measure function. NSAID=non-steroidal anti-inflammatory drug. pts=points. sig=significant. TENS=transcutaneous nerve stimulation.

*Modified Jadad quality score (listed first) (range 0-5); Cochrane Back Review Group quality score (0-10). Higher alternative score indicates one outcome was assessed masked.

†Measurement points: follow up time-points are duration from end of treatment

Table 4. Characteristics of Chinese acupuncture studies included in meta-analysis

Reference, Date	**Quality***	**Acup style†**; no. of sessions; times per week	**N**	**Control** Intervention(s)	**N**	**Any co-interventions** (except drugs) in all groups	**Pain** measure	**Function** measure	**Global score, work or drug intake**	**Measurement points ‡** (not all outcomes)
Carlsson & Sjolund, 2001[24]	4; 5	Formula EA; 10; 1	34	Sham TENS	16	-	VAS morning	-	Global 2 pts (masked evaluator); fit for work; tablet intake	eot, 1m, 3m, 6m
Cherkin et al, 2001[25]	3; 7	Individual EA; 8; 1	94	a)Massage	78	Exercise (some patients)	Bothersomeness rating	Roland Disability modified	Analgesic use	eot, 1y
	3; 7			b)Self-care education	90					
Coan et al, 1980[26]	2; 2	Individual EA; 11; ?	25	No treatment	25	-	Numerical rating || **	Activity limitation (4 pts) ||	Global 2 pts (patient)	1m
Edelist et al, 1976[28]	2; 3	Formula EA; 3; 3	15	Sham acup	15	-	-	-	Global 2 pts (patient)	eot
Giles & Muller, 1999[30]	3; 2	Flexible formula EA; 6; 2	18	a)Manipulation	32	-	VAS || **	Oswestry || **	Global 2 pts (patient)	eot
	3; 2			b)Analgesic medication	20					
Giles & Muller, 2003[31]	3; 4	Individual; 18; 2	36	a)Analgesic medication	43	-	VAS || **	Oswestry || **	Global ('recovery')	eot
				b)Manipulation	36					eot
Grant et al, 1999[32]	3; 6	Individual; 8; 2	32	TENS	28	-	VAS || **	-	[Analgesic use] §	eot, 3m
Ito, 2000[35]	2; 3	Individual EA; 4; 2	14	Medication: drugs & herbs	12	-	Numerical rating scale	JOA		eot
Kerr, 2003[36]	2; 2	Formula; 6; 1	30	Sham TENS	30	Exercise booklet	VAS	-	Global 2 pts (patient)	eot, 6m
Lehmann et al, 1986[40]	2; 2	Individual EA; 6; 2	17	Sham TENS	18	Multidisciplinary education	VAS ||	-	Fit for work	eot, 6m
	2; 2			TENS	18					
Leibing et al, 2002[41]	4; 6	Formula; 20; 5	50	a)Sham acup	50	Physical therapy	VAS	Pain Disability Index		eot, 9m
	3; 5			b)No treatment	50					

Reference, Date	**Quality***	**Acup style†**; no. of sessions; times per week	**N**	**Control** Intervention(s)	**N**	**Any co-interventions** (except drugs) in all groups	**Pain** measure	**Function** measure	**Global score, work or drug intake**	**Measurement points ‡** (not all outcomes)
Mazieres et al, 1985[15]	1; 4	?; 6; 3	17	No treatment	17	Physical therapy	VAS ||	-		eot
Mendelson et al, 1983[43]	3; 3	Flexible formula; 8; 2	36	Sham acup	41	-	VAS	-		eot, 10w
Meng et al, 2003[44]	3; 5	Flexible formula EA; 10; 2	31	No treatment	24	-	VAS	Roland Disability	Global 2 pts (patient)	[1w], 4w
Molsberger et al, 2002[45]	4; 6	Flexible formula; 12; 3	65	a)Sham acup	61	Conventional orthopedic treatment	VAS	-	Global 4 pts (patient); analgesic use] §	eot, 3m
	3; 4			b)No treatment	60					
Nobili et al, 1985[46]	2; 4	Formula EA; 7; 3	24	TENS	24	-	50% reduction in VAS ¶	-		[eot], 1m, [3m], 6m
Sakai et al, 2001[47]	3; 4	Flexible formula EA; 5; 3	31	TENS	33	Poultices	VAS	JOA		eot
Thomas & Lundeberg, 1994[48]	2; 3	Flexible formula EA; 7; 2	33	No treatment	10	-	Number of descriptors checked	Listed activities with <50% pain	Global 3 pt (patient)	eot, 6m
von Mencke et al, 1988[56]	3; 6	Flexible formula; 6; 2	35	Sham acup	30	-	Pain scale (undefined)	-		eot
Yeung et al, 2003[52]	2; 6	Formula EA; 12; 3	26	No treatment	26	Back exercises	Numerical rating scale	Aberdeen disability	Analgesic use	[eot], 1m, 3m
Zhang DW et al, 2002[54]	1; 1	Individual EA; 10 to 60; 5	96	No treatment	98	Tuina massage	-	-	Global 3 pts (patient)	eot
Zhang ZY, 2002[55]	1; 1	Individual EA; 20; 2	30	No treatment	31	Tuina massage, cupping	-	-	Global 3 pts (patient)	eot

Abbreviations: acup=acupuncture. EA=electroacupuncture. eot=end of treatment. JOA=Japanese Orthopaedic Association measure function. NS=Not significant. pts=points. VAS=visual analogue scale.

* Modified Jadad quality score (listed first) (range 0-5; ≥3 regarded as indicating good quality); Cochrane Back Review Group quality score (0-10; ≥5 regarded as indicating good quality); † EA indicates the report states that needles were stimulated electrically in some or all patients; ‡ Measurement points: follow up time-points are duration from end of treatment; § Outcomes or time-points in square brackets were not combined in meta-analysis; || Standard deviation imputed from measure's range; ¶ Means imputed from percentage with pain relief; ** Means imputed from medians.

Data synthesis

Figure 1 shows the short-term effectiveness of acupuncture on pain compared with each of the different controls. Acupuncture is significantly more effective than both sham acupuncture, sham TENS, and no additional treatment controls for chronic low back pain patients. The effect size of -0.58 for the comparison of acupuncture versus sham acupuncture corresponds with a clinically

important improvement of 14.5 mm on the VAS scale. Acupuncture is not more effective than other active treatments, and was significantly less effective than spinal manipulation.

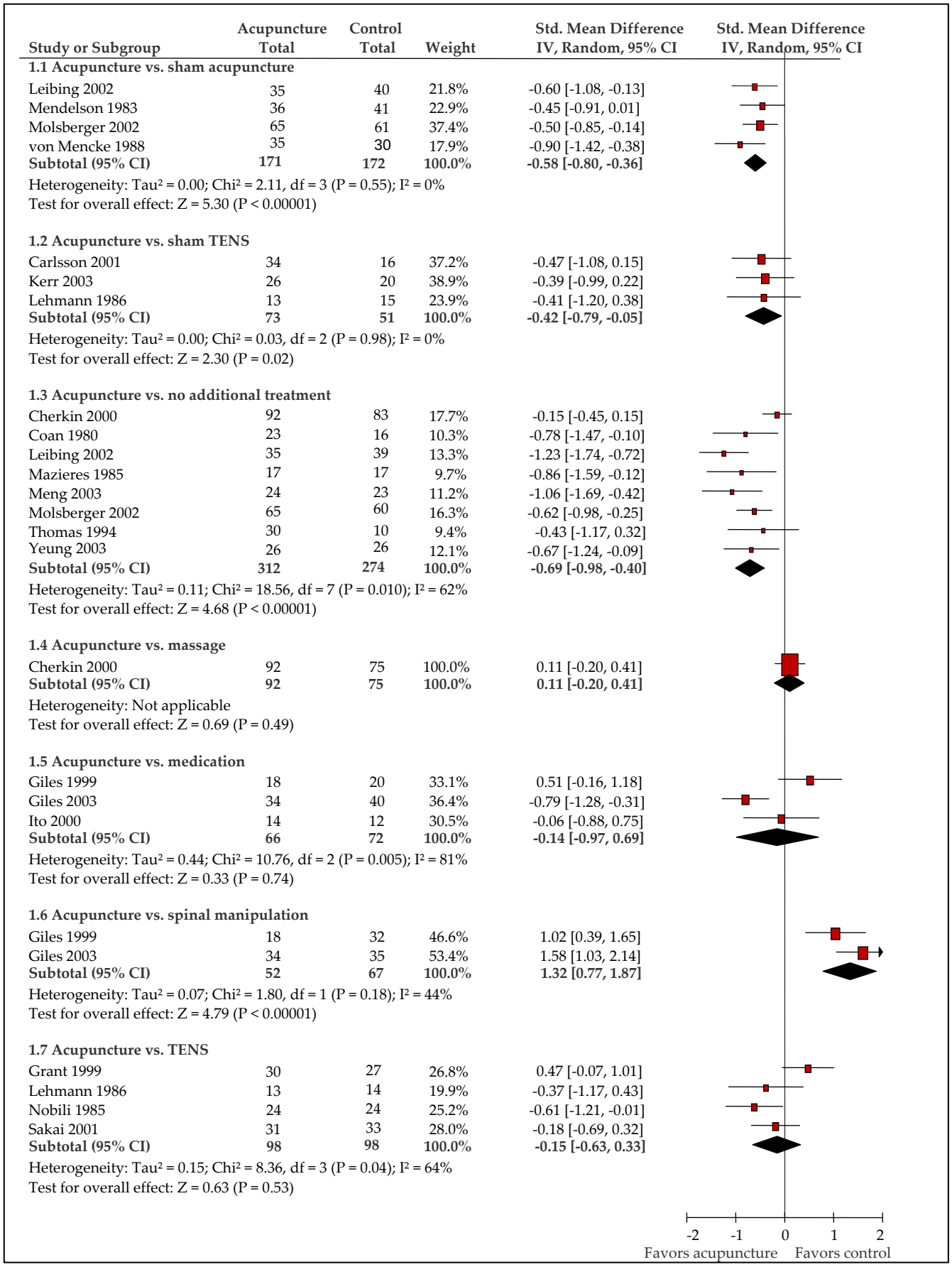


Figure 1. Short term effects of Chinese-style acupuncture on chronic pain

Figure 2 shows the long-term effects of acupuncture on pain. Acupuncture is significantly more effective than both no additional treatment and sham TENS. Acupuncture is also more effective than the four sham controlled trials combined together (standardized mean difference -0.61 [95% CI, -0.21 to -1.01]). Acupuncture is significantly worse than massage (based on the results of one trial).

Figures 3 and 4 show the short- and long-term effects of acupuncture, compared separately with sham, no additional treatment, and the various active controls, on the outcomes of functional status (Figure 3) and overall improvement (Figure 4). Since sham acupuncture and sham TENS trials were statistically homogeneous, the analyses for Figures 3 and 4 were performed after combining these studies.

For improving functioning, acupuncture was significantly more effective than only the no additional treatment control, and only in the short-term (-0.62 [95% CI, -0.30 to -0.95]) (see Figure 3).

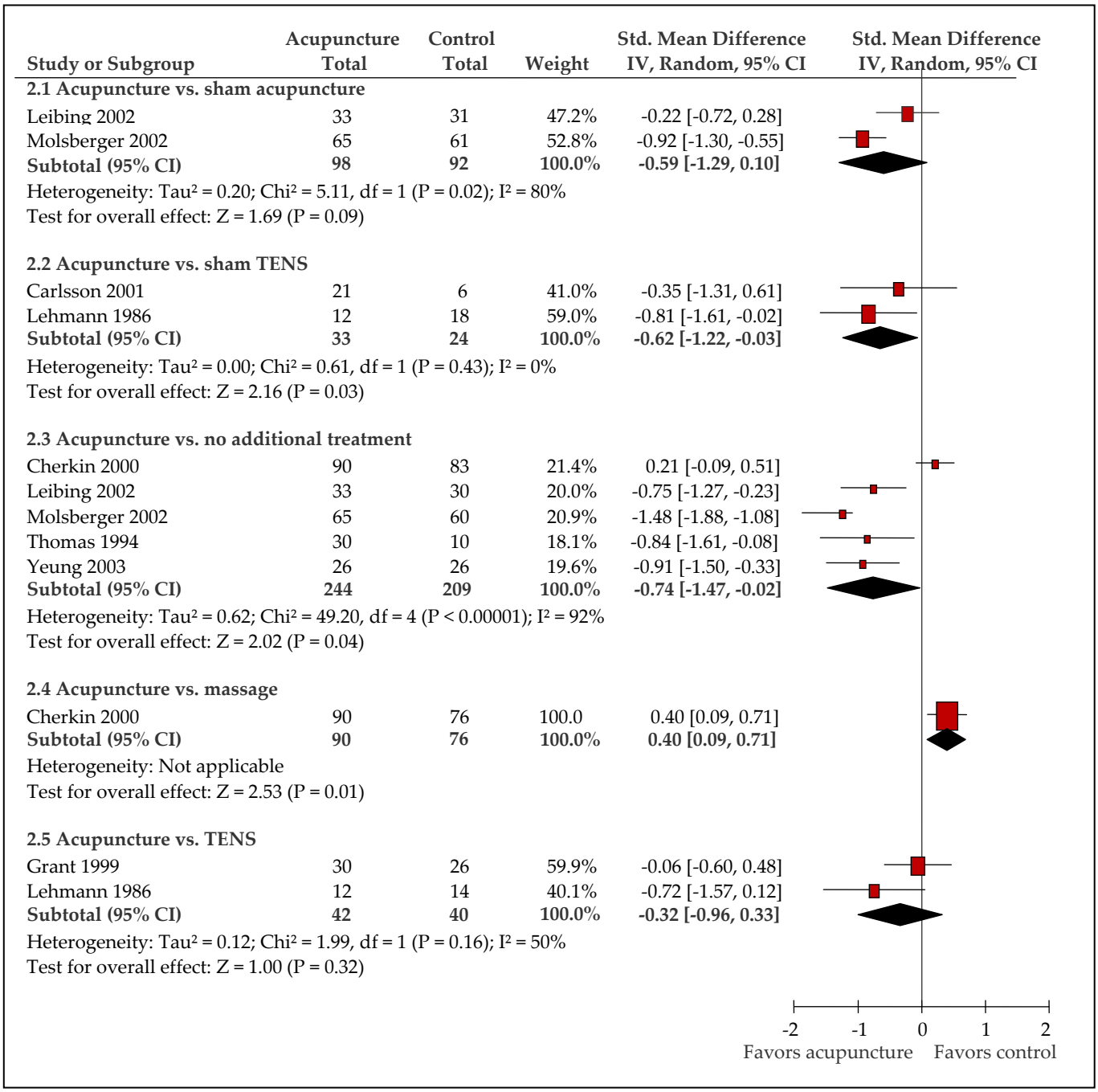


Figure 2. Long term effects of Chinese-style acupuncture on chronic pain

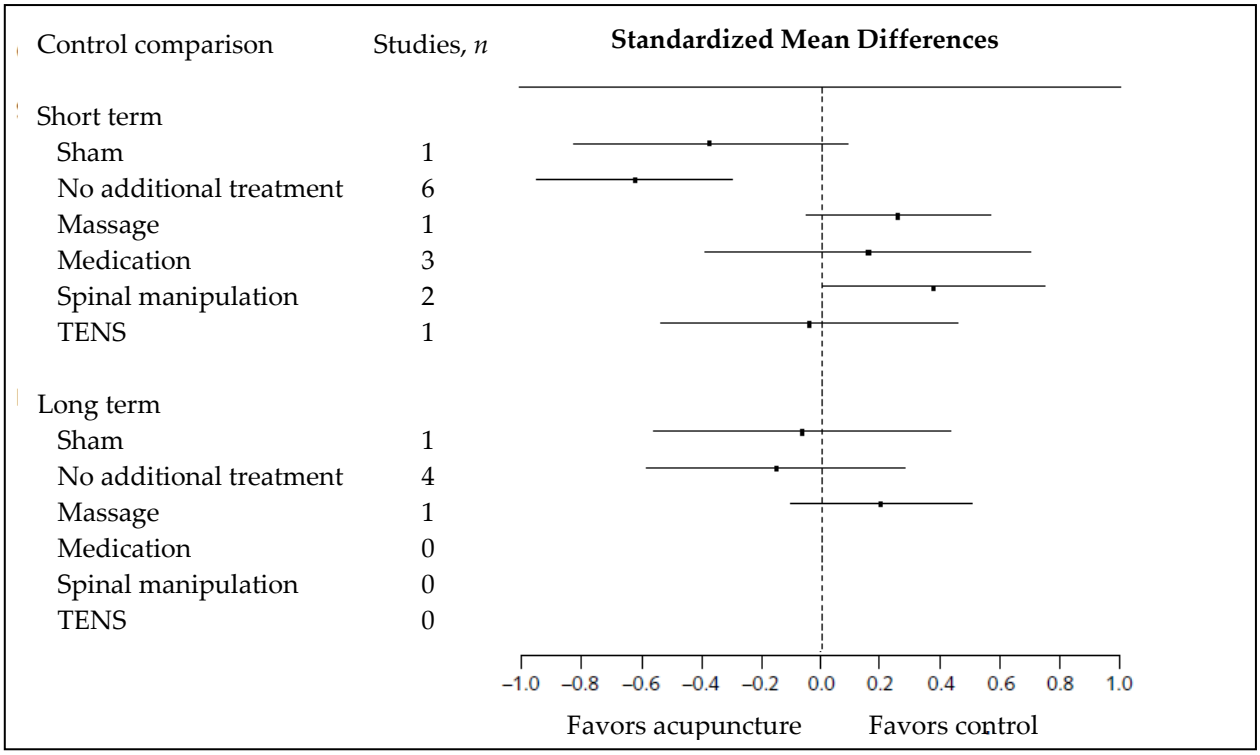


Figure 3. Functional status

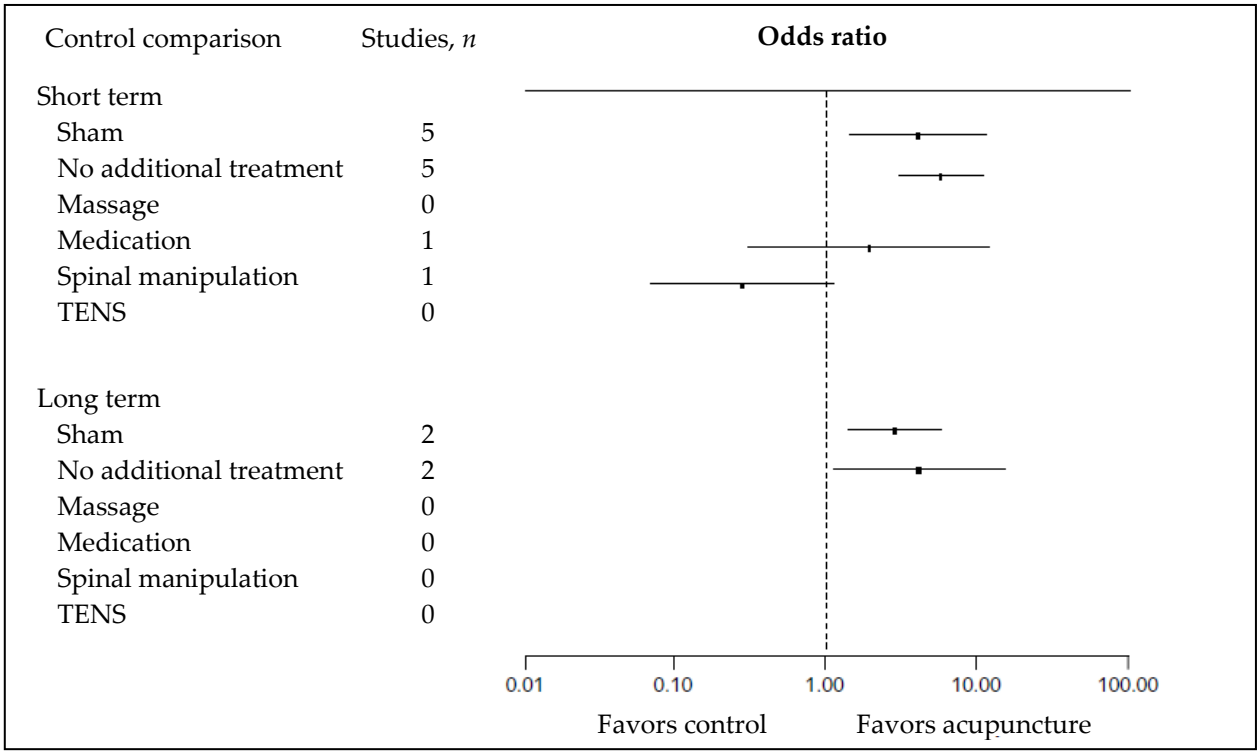


Figure 4. Overall improvement

For overall improvement, acupuncture was significantly more effective than sham controls and no additional treatment in both the short- and long- term (see Figure 4). There were only two trials that assessed return to work and three that assessed analgesic use and from these data no conclusions can be drawn (see Supplementary Appendix Figures 2 and 3).

For the seven sham controlled trials, the results were clearly homogeneous (see Figure 1) and there were no positive heterogeneity tests on any criteria of either of the quality scales or on any patient- or treatment-related factor. For the no additional treatment controlled trials, there were three positive heterogeneity tests, which were all due entirely to a heterogeneous factor in the Cherkin et al trial, which had different results from all the other no additional treatment controlled trials. None of the following sensitivity analyses affected the statistical significance of the results: adjusting the values of the imputed means and standard deviations; excluding the trials not reporting randomization procedures; or including only trials that included co-interventions or only trials that excluded co-interventions.

The interpretation of our funnel plots was hindered by the small number of large trials, and the small number of trials overall (see Supplementary Appendix Figures 4-6).

DISCUSSION

Overall, data from the 22 randomized trials included in the meta-analyses suggest that acupuncture is an effective treatment for chronic low back pain. For patients with acute low back pain, data are sparse and inconclusive. There is currently no evidence that acupuncture is more effective than other active therapies for chronic low back pain, and acupuncture appears less effective than spinal manipulation on the strength of two studies.

Drawing generalized conclusions is complicated by the fact that different RCTs used different controls, included patients with different categories of pain, and evaluated different types of outcomes. To address the heterogeneity among studies, our protocol stipulated that short-term pain relief would be the primary outcome and that multiple meta-analyses would be conducted, stratifying by type of acupuncture, control, pain, and outcome. For the primary outcome of short-term pain relief among chronic low back pain patients, acupuncture demonstrated statistically significant and clinically important benefits when compared with a sham treatment or no additional treatment, but not when compared with other active therapies.

Studies comparing acupuncture with various different control types were analyzed separately because different control comparators address different questions. In addition, each control has its advantages and limitations, which must be considered in interpreting the analysis results. A sham acupuncture control is intended to control for the non-specific effects of acupuncture and addresses the following question: Are acupuncture's effects due to the specific effects of stimulating the points needled or to the non-specific effects of the needles and/or the treatment milieu in which acupuncture is administered? Sham acupuncture, while considered the most rigorous control because it enables the blinding of patients and/or evaluators to treatment received, may also have the undesirable effect of resulting in an underestimation of the specific effects of acupuncture. Because the sham needles have the potential to unintentionally stimulate a physiological response, sham acupuncture could have some specific analgesic effects[57,58] especially when needles that penetrate the skin are used, as was the case for all sham acupuncture RCTs included in our meta-analysis. The sham-controlled RCTs generally showed less of a benefit of acupuncture compared with the no-treatment-controlled RCTs (see Figures 1 and 2), as would be expected, assuming needling at non-specific points has some therapeutic benefits, or assuming that acupuncture is associated with a placebo response. The no additional treatment control, while not blinded, and therefore considered less methodologically rigorous, may be better than the sham control for estimating the total specific plus non-specific effects of acupuncture under real-life conditions, as opposed to the de-contextualized effects of needling specific points, best measured using the sham control. Meta-analytic results from the no additional treatment control RCTs address the question of whether a policy of administering acupuncture would be preferable to a policy of doing nothing for relieving the pain of low back pain patients. The active control studies address the question of

whether acupuncture is better than other unspecified active treatments for low back pain patients. There was no evidence that acupuncture was better than any other active treatment.

Our positive results concur with those from a meta-analysis published by two of the authors (EE and AW) in 1998, in which the primary analysis showed an overall odds ratio of improvement of 2.30 (95% CI 1.28 to 4.13) in favor of acupuncture.[4] In this earlier review all the identified studies – regardless of type of control, type of pain, type of outcome, or type of acupuncture – were grouped together for the same primary meta-analysis, thus allowing the authors to draw only general conclusions about acupuncture, as compared to an unspecified control for overall short-term improvement. Because the present meta-analysis includes more than twice as many RCTs than the one from 1998,[4] including RCTs published since 1998 as well as earlier RCTs that were not identified for the previous meta-analysis,[4] our power to compare acupuncture versus distinct control categories is improved. While the 1998 review[4] could only suggest that acupuncture may be superior to a sham control, this update provides stronger evidence for acupuncture's superiority to sham. We were also now able to examine quantitatively the longer term effects of acupuncture. The results suggest that acupuncture not only provides short-term amelioration but may also provide longer term relief from chronic low back pain; however, additional sham controlled RCTs are necessary to verify that these longer term benefits are a specific effect of acupuncture.

Our results differ from a 1999 Cochrane Review of acupuncture for low back pain that did not conclude that acupuncture was effective.[6] One suspected reason is that five high quality, relatively large, sham- or no-treatment-control RCTs[24,41,44,45,52] have been published since 1999. Four of these RCTs significantly favoured acupuncture. The next revision of the Cochrane Review will likely take account of these new RCTs. Another possible explanation for the difference in the findings is that we used a different data analysis approach than does the Cochrane Review. While we used meta-analysis, the Cochrane reviewers declined to combine the trials because of trial heterogeneity. A potential problem of any strictly qualitative approach is that it can be more subjective than meta-analysis: reviewer bias can potentially influence the reviewers' classification of studies as positive or negative, as well as how individual studies impact the review's overall conclusions.[59] The homogeneity of results among both the no additional treatment controlled RCTs (with the exception of the Cherkin et al study) and the sham-controlled RCTs strengthens our confidence both in the appropriateness of using meta-analysis in this systematic review and in the results of the meta-analyses.

Our search for eligible reports was nearly comprehensive, resulting in the identification of 33 RCTs--more than twice as many as included in any previous review on this topic. Because of resource limitations, we did not search Chinese databases. The likely consequence of omitting Chinese RCTs would be a deflation of our estimates of effects, as acupuncture RCTs published in China have been shown to be positive in 100% of all cases.[60] The comprehensiveness of our searches (with the exception of Chinese databases) might be expected to limit the effects of publication bias.

Current data show that acupuncture is more effective than sham acupuncture and no additional treatment for providing short-term relief of chronic low back pain. This short-term relief appears to be sustained over the longer term; however, it is still too early to be certain of a sustained effect because the longer-term follow-up data that currently exist are limited in quantity and quality. The main results of two large RCTs in the United Kingdom[61] and Germany,[62] which remain unpublished but have been presented at conferences, are in line with the results of our meta-analysis. More research is needed to evaluate acupuncture's effects on acute low back pain and the evidence comparing acupuncture to other active treatments is inconclusive. Though current estimates of acupuncture's effects on chronic low back pain are statistically significant and clinically important, they are still somewhat preliminary, and the publication of several large ongoing trials will have a major impact on the evidence picture.

REFERENCES

1. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet* 1999;354(9178):581-5.
2. Wolsko PM, Eisenberg DM, Davis RB, Kessler R, Phillips RS. Patterns and perceptions of care for treatment of back and neck pain: results of a national survey. *Spine (Phila Pa 1976)* 2003;28(3):292-7; discussion 98.
3. Cherkin DC, Deyo RA, Sherman KJ, Hart LG, Street JH, Hrbek A, Davis RB, Cramer E, Milliman B, Booker J, et al. Characteristics of visits to licensed acupuncturists, chiropractors, massage therapists, and naturopathic physicians. *J Am Board Fam Pract* 2002;15(6):463-72.
4. Ernst E, White AR. Acupuncture for back pain: a meta-analysis of randomized controlled trials. *Arch Intern Med* 1998;158(20):2235-41.
5. Smith LA, Oldman AD, McQuay HJ, Moore RA. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. *Pain* 2000;86(1-2):119-32.
6. Tulder MV, Cherkin DC, Berman B, Lao L, Koes BW. Acupuncture for low back pain. *Cochrane Database Syst Rev* 2000(2):CD001351.
7. Cherkin DC, Sherman KJ, Deyo RA, Shekelle PG. A review of the evidence for the effectiveness, safety, and cost of acupuncture, massage therapy, and spinal manipulation for back pain. *Ann Intern Med* 2003;138(11):898-906.
8. Tsukayama H, Yamashita H. Systematic review of clinical trials on acupuncture in the Japanese literature. *Clinical Acupuncture and Oriental Medicine* 2002;3:105-13.
9. Manheimer E, Lefebvre C, Dickersin K (editors). Chapter 5: Locating and selecting studies. In: Alderson P, Green S, Higgins J (editors). *Cochrane Reviewer's Handbook 4.2.1* [updated December 2003] Chichester, UK: John Wiley and Sons Ltd, 2003.
10. Birch S. Systematic reviews of acupuncture - are there problems with these? *Clinical Acupuncture and Oriental Medicine* 2001;2:17-22.
11. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12.
12. van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. *Spine (Phila Pa 1976)* 1997;22(20):2323-30.
13. Assendelft WJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG. Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Ann Intern Med* 2003;138(11):871-81.
14. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21(11):1539-58.
15. Mazières B, Frize B, Bayourthe L, Guiraud G. Acupuncture treatment of chronic low back pain: a short-term controlled trial [Abstract]. *In: Abstracts of the World Congress of Rheumatology, Sydney, Australia, 19–25 May 1985. International League of Associations for Rheumatology* 1985;100:F53.
16. Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;86(3):217-25.
17. Birch S. Issues to consider in determining an adequate treatment in a clinical trial of acupuncture. *Complement Ther Med* 1997;5:8-12.
18. Hogeboom CJ, Sherman KJ, Cherkin DC. Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists. *Complement Ther Med* 2001;9(3):154-66.
19. Sherman KJ, Hogeboom CJ, Cherkin DC. How traditional Chinese medicine acupuncturists would diagnose and treat chronic low back pain: results of a survey of licensed acupuncturists in Washington State. *Complement Ther Med* 2001;9(3):146-53.
20. Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine (Phila Pa 1976)* 2001;26(13):1418-24.
21. Yue S. Acupuncture for chronic back and neck pain. *Acupunct Electrother Res* 1978;3:323-24.
22. Gallacchi G, Muller W, Plattner GR, Schnorrenberger CC. [Acupuncture and laser treatment in cervical and lumbar syndrome]. *Schweiz Med Wochenschr* 1981;111(37):1360-6.
23. Lopacz S, Gralewski Z. [Evaluation of the results of treatment of low backache by acupuncture or suggesting (preliminary report)]. *Neurol Neurochir Pol* 1979;13(4):405-9.
24. Carlsson CP, Sjolund BH. Acupuncture for chronic low back pain: a randomized placebo-controlled study with long-term follow-up. *Clin J Pain* 2001;17(4):296-305.
25. Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, Deyo RA. Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. *Arch Intern Med* 2001;161(8):1081-8.
26. Coan RM, Wong G, Ku SL, Chan YC, Wang L, Ozer FT, Coan PL. The acupuncture treatment of low back pain: a randomized controlled study. *Am J Chin Med* 1980;8(1-2):181-9.
27. Duplan B, Cabanel G, Piton JL, Grauer JL, Phelip X. [Acupuncture and sciatica in the acute phase. Double-blind study of 30 cases]. *Sem Hop* 1983;59(45):3109-14.
28. Edelist G, Gross AE, Langer F. Treatment of low back pain with acupuncture. *Can Anaesth Soc J* 1976;23(3):303-6.

29. Garvey TA, Marks MR, Wiesel SW. A prospective, randomized, double-blind evaluation of trigger-point injection therapy for low-back pain. *Spine (Phila Pa 1976)* 1989;14(9):962-4.
30. Giles LG, Muller R. Chronic spinal pain syndromes: a clinical pilot trial comparing acupuncture, a nonsteroidal anti-inflammatory drug, and spinal manipulation. *J Manipulative Physiol Ther* 1999;22(6):376-81.
31. Giles LG, Muller R. Chronic spinal pain: a randomized clinical trial comparing medication, acupuncture, and spinal manipulation. *Spine (Phila Pa 1976)* 2003;28(14):1490-502; discussion 502-3.
32. Grant DJ, Bishop-Miller J, Winchester DM, Anderson M, Faulkner S. A randomized comparative trial of acupuncture versus transcutaneous electrical nerve stimulation for chronic back pain in the elderly. *Pain* 1999;82(1):9-13.
33. Guerreiro da Silva JB, Nakamura MU, Cordeiro JA, Kulay L, Jr. Acupuncture for low back pain in pregnancy--a prospective, quasi-randomised, controlled study. *Acupunct Med* 2004;22(2):60-7.
34. Gunn CC, Milbrandt WE, Little AS, Mason KE. Dry needling of muscle motor points for chronic low-back pain: a randomized clinical trial with long-term follow-up. *Spine (Phila Pa 1976)* 1980;5(3):279-91.
35. Ito K. Controlled comparative trial of acupuncture treatment for lumbago. Report for Foundation for Training and Licensure Examination in Annma, Massage,Acupressure, Acupuncture and Moxibustion. Tokyo, Japan: Toyo Ryoho Kenshu Shiken Zaidan (Foundations for Training and Licensure Examination in Annma—Massage, Acupressure, Acupuncture, and Moxibustion); 2000.
36. Kerr DP, Walsh DM, Baxter D. Acupuncture in the management of chronic low back pain: a blinded randomized controlled trial. *Clin J Pain* 2003;19(6):364-70.
37. Kittang G, Melvaer T, Baerheim A. [Acupuncture contra antiphlogistics in acute lumbago]. *Tidsskr Nor Laegeforen* 2001;121(10):1207-10.
38. Kurosu Y. Acupuncture and moxibustion for lumbago (II)—comparative experiment of the therapeutic effectiveness of acupuncture and garlic moxibustion. *Nippon Gan Chiryo Gakkai Shi* 1979;28:31-4.
39. Kvorning N, Holmberg C, Grennert L, Aberg A, Akeson J. Acupuncture relieves pelvic and low-back pain in late pregnancy. *Acta Obstet Gynecol Scand* 2004;83(3):246-50.
40. Lehmann TR, Russell DW, Spratt KF, Colby H, Liu YK, Fairchild ML, Christensen S. Efficacy of electroacupuncture and TENS in the rehabilitation of chronic low back pain patients. *Pain* 1986;26(3):277-90.
41. Leibing E, Leonhardt U, Koster G, Goerlitz A, Rosenfeldt JA, Hilgers R, Ramadori G. Acupuncture treatment of chronic low-back pain -- a randomized, blinded, placebo-controlled trial with 9-month follow-up. *Pain* 2002;96(1-2):189-96.
42. Macdonald AJ, Macrae KD, Master BR, Rubin AP. Superficial acupuncture in the relief of chronic low back pain. *Ann R Coll Surg Engl* 1983;65(1):44-6.
43. Mendelson G, Selwood TS, Kranz H, Loh TS, Kidson MA, Scott DS. Acupuncture treatment of chronic back pain. A double-blind placebo-controlled trial. *Am J Med* 1983;74(1):49-55.
44. Meng CF, Wang D, Ngeow J, Lao L, Peterson M, Paget S. Acupuncture for chronic low back pain in older patients: a randomized, controlled trial. *Rheumatology (Oxford)* 2003;42(12):1508-17.
45. Molsberger AF, Mau J, Pawelec DB, Winkler J. Does acupuncture improve the orthopedic management of chronic low back pain--a randomized, blinded, controlled trial with 3 months follow up. *Pain* 2002;99(3):579-87.
46. Nobili C, Colombi A, Ambrosino G. Confronto a distanza dell'efficacia terapeutica dell'agopuntura e della TENS nel trattamento del low-back pain. *Minerva Riflessoterapeutica e Laserterapeutica* 1985;2:61-7.
47. Sakai T, Tsutani K, Tsukayama H, Nakamura T, Ikeuchi R, Kawamoto M, et al. Multi-center randomized, controlled trial of acupuncture with electric stimulation and acupuncture-like transcutaneous electrical nerve stimulation for lumbago. *Journal of Japan Society of Acupuncture and Moxibustion* 2001;51:175-84.
48. Thomas M, Lundberg T. Importance of modes of acupuncture in the treatment of chronic nociceptive low back pain. *Acta Anaesthesiol Scand* 1994;38(1):63-9.
49. Tsukayama H, Yamashita H, Amagai H, Tanno Y. Randomised controlled trial comparing the effectiveness of electroacupuncture and TENS for low back pain: a preliminary study for a pragmatic trial. *Acupunct Med* 2002;20(4):175-80.
50. von Mencke M, Wieden T, Hoppé M, Pörschke W, Hoffmann O, Herget H. Akupunktur des Schulter-Arm-Syndroms und der Lumbalgie/Ischialgie - Zwei Prospektive Doppelblind-Studien (Teil 2). *Akupunktur Theorie und Praxis* 1989;5:5-13.
51. Wedenberg K, Moen B, Norling A. A prospective randomized study comparing acupuncture with physiotherapy for low-back and pelvic pain in pregnancy. *Acta Obstet Gynecol Scand* 2000;79(5):331-5.
52. Yeung CK, Leung MC, Chow DH. The use of electro-acupuncture in conjunction with exercise for the treatment of chronic low-back pain. *J Altern Complement Med* 2003;9(4):479-90.
53. Yokoyama M, Sun X, Oku S, Taga N, Sato K, Mizobuchi S, Takahashi T, Morita K. Comparison of percutaneous electrical nerve stimulation with transcutaneous electrical nerve stimulation for long-term pain relief in patients with chronic low back pain. *Anesth Analg* 2004;98(6):1552-6, table of contents.
54. Zhang D, Liang X, Wang Q. [Clinical research on the treatment of lumbar intervertebral disc protrusion by the combination of acupuncture and massage]. *Shanghai Journal of Acupuncture and Moxibustion* 2002;21:22-3.

55. Zhang Z. [Observations on the curative effect of acupuncture and massage on lumbar intervertebral disc protrusion]. *Shanghai Journal of Acupuncture and Moxibustion* 2002;21:15-6.
56. von Mencke M, Wieden T, Hoppé M, Pörschke W, Hoffmann O, Herget H. Akupunktur des Schulter-Arm-Syndroms und der Lumbalgie/Ischialgie - Zwei Prospektive Doppelblind-Studien (Teil 1). *Akupunktur Theorie und Praxis* 1988;4:204-15.
57. Le Bars D, Dickenson AH, Besson JM. Diffuse noxious inhibitory controls (DNIC). I. Effects on dorsal horn convergent neurones in the rat. *Pain* 1979;6(3):283-304.
58. Lewith GT, Vincent C. On the evaluation of the clinical effects of acupuncture: a problem reassessed and a framework for future research. *J Altern Complement Med* 1996;2(1):79-90; discussion 91-100.
59. Deeks J, Higgins J, Altman D (editors). Chapter 8: Analysing and presenting results. In: Alderson P, Green S, Higgins J, editors. *Cochrane Reviewers' Handbook 4.2.1 [updated December 2003] Section 8. In: The Cochrane Library, Issue 1*. Chichester, UK: John Wiley and Sons Ltd, 2004.
60. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;19(2):159-66.
61. Thomas KJ, Fitter M, Brazier J, MacPherson H, Campbell M, Nicholl JP, Roman M. Longer-term clinical and economic benefits of offering acupuncture to patients with chronic low back pain assessed as suitable for primary care management. *Complement Ther Med* 1999;7(2):91-100.
62. Brinkhaus B, Becker-Witt C, Jena S, Linde K, Streng A, Wagenpfeil S, Irnich D, Hummelsberger J, Melchart D, Willich SN. Acupuncture Randomized Trials (ART) in patients with chronic low back pain and osteoarthritis of the knee - design and protocols. *Forsch Komplementarmed Klass Naturheilkd* 2003;10(4):185-91.

Chapter 8. Acupuncture for peripheral joint osteoarthritis: Cochrane review

Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland LS, van der Windt DA, Berman BM, Bouter LM. Acupuncture for peripheral joint osteoarthritis. Cochrane Database Syst Rev 2010(1):CD001977.

ABSTRACT

BACKGROUND: Peripheral joint osteoarthritis is a major cause of pain and functional limitation. Few treatments are safe and effective.

OBJECTIVES: To assess the effects of acupuncture for treating peripheral joint osteoarthritis.

SEARCH STRATEGY: We searched the Cochrane Central Register of Controlled Trials (*The Cochrane Library* 2008, Issue 1), MEDLINE, and EMBASE (both through December 2007), and scanned reference lists of articles.

SELECTION CRITERIA: Randomized controlled trials (RCTs) comparing needle acupuncture with a sham, another active treatment, or a waiting list control group in people with osteoarthritis of the knee, hip, or hand.

DATA COLLECTION AND ANALYSIS: Two authors independently assessed trial quality and extracted data. We contacted study authors for additional information. We calculated standardized mean differences using the differences in improvements between groups.

MAIN RESULTS: Sixteen trials involving 3498 people were included. Twelve of the RCTs included only people with OA of the knee, three only OA of the hip, and one a mix of people with OA of the hip and/or knee. In comparison with a sham control, acupuncture showed statistically significant, short-term improvements in osteoarthritis pain (standardized mean difference -0.28, 95% confidence interval -0.45 to -0.11; 0.9 point greater improvement than sham on 20 point scale; absolute percent change 4.59%; relative percent change 10.32%; 9 trials; 1835 participants) and function (-0.28, -0.46 to -0.09; 2.7 point greater improvement on 68 point scale; absolute percent change 3.97%; relative percent change 8.63%); however, these pooled short-term benefits did not meet our predefined thresholds for clinical relevance (i.e., 1.3 points for pain; 3.57 points for function) and there was substantial statistical heterogeneity. Additionally, restriction to sham-controlled trials using shams judged most likely to adequately blind participants to treatment assignment (which were also the same shams judged most likely to have physiological activity), reduced heterogeneity and resulted in pooled short-term benefits of acupuncture that were smaller and non-significant. In comparison with sham acupuncture at the six-month follow-up, acupuncture showed borderline statistically significant, clinically irrelevant improvements in osteoarthritis pain (-0.10, -0.21 to 0.01; 0.4 point greater improvement than sham on 20 point scale; absolute percent change 1.81%; relative percent change 4.06%; 4 trials; 1399 participants) and function (-0.11, -0.22 to 0.00; 1.2 point greater improvement than sham on 68 point scale; absolute percent change 1.79%; relative percent change 3.89%). In a secondary analysis versus a waiting list control, acupuncture was associated with statistically significant, clinically relevant short-term improvements in osteoarthritis pain (-0.96, -1.19 to -0.72; 14.5 point greater improvement than waiting list on 100 point scale; absolute percent change 14.5%; relative percent change 29.14%; 4 trials; 884 participants) and function (-0.89, -1.18 to -0.60; 13.0 point greater improvement than waiting list on 100 point scale; absolute percent change 13.0%; relative percent change 25.21%). In the head-on comparisons of acupuncture with the 'supervised osteoarthritis education' and the 'physician consultation' control groups, acupuncture was associated with clinically relevant short-

and long-term improvements in pain and function. In the head on comparisons of acupuncture with 'home exercises/advice leaflet' and 'supervised exercise', acupuncture was associated with similar treatment effects as the controls. Acupuncture as an adjuvant to an exercise based physiotherapy program did not result in any greater improvements than the exercise program alone. Information on safety was reported in only eight trials and even in these trials there was limited reporting and heterogeneous methods.

AUTHORS' CONCLUSIONS: Sham-controlled trials show statistically significant benefits; however, these benefits are small, do not meet our pre-defined thresholds for clinical relevance, and are probably due at least partially to placebo effects from incomplete blinding. Waiting list-controlled trials of acupuncture for peripheral joint osteoarthritis suggest statistically significant and clinically relevant benefits, much of which may be due to expectation or placebo effects.

The online version of the thesis Table of Contents includes a link to the full-text of the following preliminary version of this review, published in *Annals of Internal Medicine* in 2007:

- Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for osteoarthritis of the knee. *Ann Intern Med*. 2007 Jun 19;146(12):868-77.

PLAIN LANGUAGE SUMMARY

Acupuncture for osteoarthritis

This summary of a Cochrane review presents what we know from research about the effect of acupuncture on osteoarthritis.

The review shows that in people with osteoarthritis,

-Acupuncture may lead to small improvements in pain and physical function after 8 weeks.

-Acupuncture may lead to small improvements in pain and physical function after 26 weeks.

-We often do not have precise information about side effects and complications. This is particularly true for rare but serious side effects. Possible side effects of acupuncture treatment include minor bruising and bleeding at the site of needle insertion.

What is osteoarthritis and what is acupuncture?

Osteoarthritis (OA) is a disease of the joints, such as your knee or hip. When the joint loses cartilage, the bone grows to try and 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 unstable. This can affect your physical function or ability to use your knee.

According to the philosophy of traditional acupuncture, energy circulates in 'meridians' located throughout the body. Pain or ill health happens when something occurs to cause this meridian energy circulation to be blocked. The way to restore health is to stimulate the appropriate combination of acupuncture points in the body by inserting very thin needles. Sometimes in painful conditions, electrical stimulation along with the needles is also used. According to acupuncture theory, one way you can tell that acupuncture is relieving pain is that you may feel numbness or tingling, called *de qi*, where the needle is inserted.

Best estimate of what happens to people with osteoarthritis who have acupuncture:

Pain after 8 weeks:

-People who had acupuncture rated their pain to be improved by about 4 points on a scale of 0 to 20.

-People who received sham acupuncture rated their pain to be improved by about 3 points on a scale of 0 to 20.

-People who received acupuncture had a 1 point greater improvement on a scale of 0 to 20 (5% absolute improvement).

Pain after 26 weeks:

-People who had acupuncture rated their pain to be improved by slightly more than 3 points on a scale of 0 to 20.

-People who received sham acupuncture rated their pain to be improved by slightly less than 3 points on a scale of 0 to 20.

-People who received acupuncture had under a 1 point greater improvement on a scale of 0-20 (2% absolute improvement).

Physical function after 8 weeks:

-People who had acupuncture rated their function to be improved by about 11 points on a scale of 0 to 68.

-People who received sham acupuncture rated their function to be improved by about 8 points on a scale of 0 to 68.
-People who received acupuncture had about a 3 point greater improvement on a scale of 0-68 (4% absolute improvement).

Physical function after 26 weeks:

-People who had acupuncture rated their function to be improved by about 11 points on a scale of 0 to 68.
-People who received sham acupuncture rated their function to be improved by about 10 points on a scale of 0 to 68.
-People who received acupuncture had about a 1 point greater improvement on a scale of 0-68 (2% absolute improvement).

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON

Acupuncture compared with sham acupuncture for peripheral joint osteoarthritis						
Patient or population: Patients with peripheral joint osteoarthritis; **Intervention:** Acupuncture; **Comparison:** Sham acupuncture						
Outcomes	**Illustrative comparative risks* (95% CI)**		**Relative percent change**	**No of Partici-pants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
	Assumed risk	**Corresponding risk**				
	Control	**Acupuncture**				
Pain (short term) WOMAC scale: 0-20 points (higher is worse pain). Follow up: 8 weeks	The mean pain (short term) in the control groups was **-2.66 points**†	The mean pain (short term) in the intervention groups was **0.92 lower** (1.48 to 0.36 lower)‡	-10.32%§	1835 (9)	++low(‖,¶)	SMD -0.28 (-0.45 to -0.11) Absolute percent difference: -4.59% (0.92 point lower on a 0-20 point scale)**
Function (short term) WOMAC scale: 0-68 points (higher is worse function). Follow up: 8 weeks	The mean function (short term) in the control groups was **-7.86 points**†	The mean function (short term) in the intervention groups was **2.70 lower** (4.44 to 0.87 lower)‡	-8.63%§	1767 (8)	++low(‖)	SMD -0.28 (-0.46 to -0.09) Absolute percent difference: -3.97% (2.70 points lower on a 0-68 point scale)**
Pain (long term) WOMAC scale: 0-20 points (higher is worse pain). Follow up: 26 weeks	The mean pain (long term) in the control groups was **-2.92 points**†	The mean pain (long term) in the intervention groups was **0.36 lower** (0.75 lower to 0.04 higher)‡	-4.06%§	1399 (4)	++++high	SMD -0.10 (-0.21 to 0.01) Absolute percent difference: -1.81% (0.36 point lower on a 0-20 point scale)**
Function (long term) WOMAC scale: 0-68 points (higher is worse function). Follow up: 26 weeks	The mean function (long term) in the control groups was **-9.94 points**†	The mean function (long term) in the intervention groups was **1.21 lower** (2.43 lower to 0 higher)‡	-3.89%§	1398 (4)	++++high	SMD -0.11 (-0.22 to 0) Absolute percent difference: -1.79% (1.22 points lower on a 0-68 point scale)**
Adverse events	See comment	See comment	Not esti-mable	-	See comment	Eight RCTs described adverse events across groups, and they found that the frequency of adverse events was similar between the acupuncture and control groups. The frequency of adverse events in the acupuncture group ranged from 0% (Sangdee 2002) to 7% (Berman

						2004). Pooling of adverse events across these RCTs was not possible because of limited reporting and heterogeneous methods. No serious adverse events were reported to be associated with acupuncture.
Side effects of acupuncture-bruising and bleeding at injection site	See comment	See comment	Not estimable	-	See comment	The frequency of minor side effects of acupuncture, primarily minor bruising and bleeding at needle insertion sites, ranged from 0% (Foster 2007) to 45% (Sangdee 2002). These frequencies varied widely because of heterogeneous and scanty reporting and different definitions of what constitutes a side effect of acupuncture versus what is an inherent part of treatment (for example, occasional bruising at needle insertion site).

*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.

†The representative trial selected for calculating the percent changes from baseline was the Berman 2004 trial because this trial was sufficiently large, and because the patient characteristics and the baseline mean and SD of the control group for this trial was most similar to, and thus most representative of, the other trials.

‡We calculated the mean difference by choosing the Berman 2004 trial as a representative study, and then calculating the difference by multiplying the SMD by the SD (standard deviation) of the mean change in the control group in this study.

§We calculated the relative percent change by multiplying the SMD by the standard deviation of change in the control group of the Berman 2004 trial, dividing the result by the baseline mean in the control group of the Berman 2004 trial, and multiplying by 100 to obtain the percent.

||We could not be certain that the shams used in three of the sham-controlled trials[1-3] were sufficiently credible in fully blinding participants to the treatment being evaluated.

¶There was statistically significant heterogeneity of effect estimates between the two substrata for the following four variables for the pain outcome: success of blinding (Yes/Not sure); likely physiological activity of sham control (Yes/No); use of electrical stimulation of needles (Yes/No); and adequate number of acupuncture sessions (Yes/No).

**We calculated the absolute percent change by multiplying the SMD by the standard deviation of change in the control group of the Berman 2004 trial, dividing the result by the number of units in the scale, and multiplying by 100 to obtain the percent.

BACKGROUND

Osteoarthritis (OA), the most common form of arthritis,[4,5] is the leading cause of disability among older adults.[6,7] Non-steroidal anti-inflammatory drugs (NSAIDS) and acetaminophen are the most commonly used pharmacological agents for treating osteoarthritis.[8,9] However, according to a recent systematic review, NSAIDs are only slightly better than placebo in providing short-term pain relief, and their effects are probably too small to be meaningful to people with OA.[10] Acetaminophen is often considered a safer alternative to NSAIDs,[9] and a recent Cochrane review[11] indeed found that acetaminophen was associated with fewer adverse effects than NSAIDs in short-term randomized controlled trials (RCTs).[12] However, this review also found that acetaminophen is modestly less effective than NSAIDs,[11] and that the clinical significance of acetaminophen is questionable,[11,12] as it results in only a 5% larger improvement from baseline in pain compared to placebo in the short-term.[11] In addition, acetaminophen may be the leading cause of acute liver failure in the United States.[13]

The most recent evidence-based treatment guidelines from the UK National Institutes of Clinical Excellence[14] and the Osteoarthritis Research International[12] suggest that OA treatment should be multidisciplinary, with non-pharmacological treatments such as education, aerobic and resistance exercises, and weight loss as the "cornerstone"[14] or "initial focus"[12] of patient management, and with consideration also given to pharmacological options such as acetaminophen when further treatment is required. In a recent systematic review of OA guidelines,[12] five of the eight guidelines that considered acupuncture recommended it as an osteoarthritis treatment modality. A very recent, authoritative, guideline for knee or hip OA[12] was developed by an international, multidisciplinary group of experts using a Delphi process. With this approach, consensus recommendations among the committee members were developed by systematically reviewing and critically appraising both meta-analyses of efficacy as well as existing guidelines. This approach was described as "evidence-driven and clinically supported". This guideline committee recommended acupuncture as one of 12 possible non-pharmacological modalities for treating osteoarthritis; however, this recommendation achieved only a 69% consensus among the guideline committee members.

Many people with OA seek out complementary and alternative medicine (CAM) therapies.[15-17] For example, according to a recent US national survey,[16] 41% of a nationally representative sample of people with arthritis have used a CAM therapy in the past year. Another recent survey found that 67% of people with OA in primary care clinics were currently using at least one type of CAM therapy,[15] with glucosamine and chondroitin being by far the most commonly used therapies.[15] Though glucosamine is used by 25% of people with OA,[15] the evidence to support its use is inconsistent,[12,18] with some recent large trials showing no benefit over a placebo.[19-22] Chondroitin is used by 18% of people with OA,[15] but a recent systematic review of large, methodologically sound trials concluded that the symptomatic benefit is "minimal or non-existent".[23] All other CAM therapies for treating OA are used far less frequently than glucosamine and chondroitin.[15] For example, acupuncture is used by only about 1%[16] to 2%[15] of people with arthritis. Even among people with OA who do use acupuncture, most do not use it specifically for treating their OA.[16]

According to the philosophy of traditional acupuncture, energy circulates in 'meridians' located throughout the body. When something occurs to cause this meridian energy circulation to be blocked, pain or ill health will result. The way to restore energy circulation, health, and balance, is to stimulate the appropriate combination of the estimated 400 traditional meridian acupuncture points in the body.[24] Additional non-meridian tender points may also be used, and electrical

stimulation of the points is also common in modern acupuncture, especially for pain-related conditions. According to acupuncture theory, one indication that acupuncture is exerting its analgesic effects is that a patient may experience a sensation of numbness or tingling, called *de qi*, at the needle insertion point.

Laboratory evidence has documented a biological basis of acupuncture analgesia. For example, animal studies provide evidence that acupuncture may simply be a particular method of stimulating the nervous system to release a range of neurotransmitters - particularly opioid peptides - which are involved in the body's own pain-suppressing mechanisms.[25] Other research suggests that the noxious stimulation of the acupuncture needles may act to suppress the nervous system pathways that are involved in the sensory and affective components of pain.[26,27] Because pain often limits a patient's activity, decreased pain may improve function. Also, basic science studies show that acupuncture suppresses inflammation[28]; any decreases in inflammation may improve physical function.

Acupuncture has been demonstrated to be a safe therapy with a very low risk of serious side effects.[29-33] A systematic review of 12 prospective studies which surveyed more than one million acupuncture treatments found that the risk of a serious adverse event from acupuncture is estimated to be 0.05 per 10 000 treatments, and 0.55 per 10 000 individual patients. Treatment guidelines[9,34,35] emphasize that treatment safety is an important consideration, especially because people with knee OA are often older, use concurrent medications, and have co-morbidities. Given its safety, the question of whether or not acupuncture is effective for treating OA of the knee therefore is highly relevant.

In everyday practice, acupuncture treatment is often individualized, continually modified to take into account changes in the patient's condition, and combined with other treatments, such as herbal medicine and mind-body exercises. In contrast, in RCTs of acupuncture, a prescribed formula of acupuncture points is generally evaluated as a sole treatment. While evaluating acupuncture as a sole, isolated treatment may not reflect everyday clinical practice, it does allow for the best estimation of the specific effects of acupuncture.

Three recent systematic reviews of RCTs have evaluated the effects of acupuncture on OA. Two of these reviews included only participants with knee OA,[36,37] while the third[38] included participants with OA of any peripheral joint. Each of these reviews included a meta-analysis which showed that acupuncture was statistically significantly superior to sham acupuncture in the short-term. This review is a substantial update of the most recently published earlier review.[36] This current review has been largely rewritten from the earlier version,[36] and includes RCTs with OA of all peripheral joints, not just the knee, as well as two large, recent, knee OA RCTs not included in any previous systematic review.

OBJECTIVES

The objective of this review was to compare the effects of traditional needle acupuncture with a sham, another active treatment, or with a waiting list control, for people with OA of the knee, hip, or hand.

METHODS

Criteria for considering studies for this review

Types of studies

We included RCTs in any language. We included only RCTs with at least six weeks of observation because trials with a shorter duration were considered irrelevant for the question of whether acupuncture is helpful for people with a chronic disease like OA. In principle, a longer trial duration would seem even more desirable; however, given the limited number of trials available we considered a minimum duration of six weeks a reasonable compromise.

Types of participants

We included only studies that concerned exclusively participants with osteoarthritis of one or more of the peripheral joints (i.e., knee, hip, and hand). Studies including participants with only OA of the spine were not included. Studies that included a mix of participants with OA of the spine and OA of the peripheral joints were included only if the results for the participants with OA of the peripheral joints were reported separately from the results of the participants with OA of the spine.

Types of interventions

We included only studies evaluating traditional acupuncture. Traditional acupuncture involves inserting needles into traditional meridian points, usually with the intention of influencing energy flow in the meridian. In traditional acupuncture, needles may also be inserted at additional tender points and electrical stimulation of the needles may be used. We excluded trials of dry needling/trigger point therapy, a therapy which rejects traditional concepts of energy and meridians, and which involves inserting needles only at unnamed tender or trigger points to stimulate nerves or muscles. We also excluded RCTs of laser acupuncture and electro-acupuncture without needle insertion because most authorities believe acupuncture involves needle insertion.[39]

The control interventions were a sham intervention, a waiting list, and another active treatment. We also included trials that compared acupuncture plus another active treatment versus that other active treatment alone. Thus, we included all pragmatic trials that compared acupuncture with any other treatments (e.g., exercise, education, medication, etc.). Because our objective was to evaluate the effects of acupuncture compared to non-acupuncture controls, we excluded RCTs in which one form of acupuncture was compared only with another form.

Types of outcome measures

At least one of the following outcome measures had to be reported: pain, function, or symptom severity.

Search methods for identification of studies

To identify RCTs, we searched the Cochrane Central Register of Controlled Trials (CENTRAL, *The Cochrane Library* 2008, Issue 1), MEDLINE (through December 2007), and EMBASE (through December 2007). For our MEDLINE and CENTRAL searches, we searched the following terms as both free-text terms as well as MeSH terms (except where indicated): (acupuncture; acupuncture therapy; auriculotherapy (free text only); electroacupuncture; moxibustion; Medicine, Oriental Traditional; Medicine, Chinese Traditional) AND (arthritis; osteoarthritis; arthralgia; joint diseases; joint pain (free-text only); chronic joint symptoms (free-text only); gonarthrosis (free-text only); osteoarthrosis (free-text only); ostoarthrosis (free-text only); degenerative arthritis (free-text only)).

We combined this search strategy with a methods filter for clinical trials.[40] For our EMBASE search, we used a modified version of the MEDLINE strategy.

All RCTs included in previous systematic reviews of acupuncture for OA[36-38,41,42] were also reconsidered for inclusion in this review. We scanned bibliographies of retrieved articles for further references. Finally, we also searched databases of ongoing trials to identify details of trials that may be relevant for future updates of this review.

Two authors (EM with either KC or KL) independently considered articles for inclusion, with disagreements resolved by discussion.

Data collection and analysis

Data extraction

One author (EM) extracted the data for all trials except the one German language trial.[43] A second author (either KC or KL) completed an independent second extraction for all trials: KC completed the second extraction for five trials[44-48] and KL completed the extractions for the remaining 11 trials. Consensus was generally achieved by discussion; in the few instances where disagreements persisted, a third author made the final decision on the extraction of the data item. We independently extracted information pertaining to quality of the methods, participants, acupuncture and control interventions, and treatment outcomes (including adverse effects). We e-mailed all RCT corresponding authors and requested that they review the information about their RCT extracted into our Characteristics of included studies table, as well as review our quality assessments of their RCT. When data reported in RCT publications was incomplete or ambiguous, we requested additional information or clarification from the corresponding authors.

We extracted the outcomes of pain, function, and symptom severity for all time points reported. When a given study reported more than one pain, function, or symptom severity measure, we gave preference to the WOMAC pain, function, and total (i.e., sum of WOMAC pain, function, and stiffness) measures, respectively, because the WOMAC has been extensively and repeatedly validated in its original English version[49-51] and also in its Spanish[52] and German[53] adaptations, each of which was used in one of the trials in our meta-analyses.[3,54] In addition, the WOMAC is the most comparable between studies because most trials in this review measured outcomes using the WOMAC scale.

The method of selecting acupuncture points was categorized as individual, fixed formula, or flexible formula. For the individual style, the practitioner is free to choose any points. For the fixed formula, the same fixed points are used for all participants. For the flexible formula, a fixed formula is used and some additional points are chosen according to the symptoms or tenderness of the patient.

For cross-over trials, we included only the data before the cross-over occurred because we considered the risk for carryover effects to be prohibitive.

Risk of bias assessment

For the risk of bias assessment, we used the new tool recommended by the Cochrane Reviewer's Handbook. As recommended, we used the following six separate criteria:

- Adequate sequence generation
- Allocation concealment
- Blinding
- Incomplete outcome data addressed (up to 3 months after randomization)

- Incomplete follow-up outcome data addressed (4 to 12 months after randomization)
- Free of selective reporting

For the "other potential threats to validity" item on the risk of bias scale, we assessed the following items: groups similar at baseline regarding the most important prognostic indicators; co-interventions avoided or similar; compliance acceptable in all groups; timing of the outcome assessment in all groups similar; and intention-to-treat analysis.

As a first step in evaluating risk of bias, we copied information relevant for making a judgment on a criterion from the original publication into a table. If available, we also entered any additional information from the study authors into this table. Two reviewers independently made a judgment whether the risk of bias for each criterion was considered low, high or unclear. A third reviewer arbitrated any disagreements.

For the blinding item on the risk of bias scale, we assigned sham-controlled trials as "Unclear" rather than "Yes" because we could not be certain that all shams were sufficiently credible in fully blinding participants to the treatment being evaluated. However, we assigned the "Yes" score to sham-controlled trials that either 1) evaluated the credibility of the sham and found the sham to be indistinguishable from true acupuncture, or 2) used needle acupuncture as the sham *and also* informed participants that two different types of acupuncture were being compared (i.e., did not inform participants that a sham treatment was involved). Some trials had both blinded sham control groups and unblinded comparison groups (i.e., waiting list or other active treatment control). In the risk of bias tables the column judgment always relates to the comparison with sham interventions. In the column description we also include the assessment for the other comparison group. As the risk of bias table does not include a "not applicable" option, the item "incomplete follow-up outcome data addressed (time point greater than three months and closest to six months after randomization)" was rated as "unclear" for trials which did not follow patients longer than three months.

For osteoarthritis trials, investigators typically measure a number of outcomes at multiple time points using various outcome measurement instruments. For the selective reporting item, we considered those trials as having a low risk of bias if they reported the results of the most relevant osteoarthritis outcomes measured (typically a measure of pain and function using the WOMAC scale) for the most relevant time points (end of treatment and, if done, follow-up), and if these results made it unlikely that authors had picked them out because they were particularly favorable or unfavorable. Trials which met all criteria, or all but one criterion, were considered to be of higher quality.

For ranking the strength and quality of the evidence for a given comparison, we used the GRADE and Summary of Findings tables recommended by The Cochrane Collaboration.

Quantitative data synthesis

We placed studies into one of four categories according to which of the following comparisons were evaluated:

1) Acupuncture versus a sham intervention;
2) Acupuncture versus a waiting list;
3) Acupuncture versus another active treatment;
4) Acupuncture plus another active treatment versus that other active treatment alone.

Trials using different active treatment comparators were analyzed separately.

Trials of acupuncture for different peripheral joints were each analyzed in a separate joint-specific meta-analysis. An additional analysis that evaluated acupuncture for OA of any peripheral joint combined trials of OA of the knee with trials of OA of other peripheral joints (i.e., hip or hand). If any trials included both knee OA and other peripheral joint OA participants and reported the outcomes separately for the participants with OA of the knee and participants with OA of the other peripheral joints, the separate outcomes from the knee, hip, or hand joint participants were included in the relevant joint-specific analysis, while the results from all included participants were included in the peripheral joint OA analysis.

The outcomes of the review were the standardized mean differences of acupuncture, as compared with each comparison group, on pain, function, and symptom severity, at both the short- and longer-term follow-up time points. For our meta-analyses, we defined the short-term outcome as the measurement point closest to eight weeks, and less than or equal to three months, following randomization. We defined the longer-term outcome as the measurement point closest to six months, and more than three months, following randomization.

Standardized mean differences were calculated using the differences in improvements between groups. We used standardized mean differences as the principal measure of effect size because the trials assessed the same outcomes but measured them in various ways (e.g., WOMAC VAS and Likert scales).

For the five RCTs that used a waiting list control group,[55-59] we excluded all outcome measurements after participants on the waiting list began acupuncture.

We pooled data using the random-effects model to account for expected heterogeneity. To evaluate heterogeneity within our categories of trials, we used I^2 tests on all outcomes meta-analyzed. If there was "considerable heterogeneity", which is defined by the Cochrane Handbook for Systematic Reviews of Interventions as an I^2 value between 75% and 100%,[60] the data were not pooled; otherwise, data were pooled. However, when studies showed "substantial heterogeneity" (i.e., $I^2 \geq 50\%$[60]), the pooled results were interpreted with caution. We also tried to determine the cause of statistically heterogeneous study results, using subgroup analyses,[61] as described below. We also conducted a sensitivity analysis using a fixed-effect analysis, to assess the robustness of our findings. Although the random-effects analysis is the preferred approach because of the large heterogeneity, we also conducted a fixed-effect as a sensitivity analysis because the larger studies are more valid and thus these studies will remain most influential in a fixed-effect analysis.

To allow for a more clinically relevant interpretation, we also evaluated whether the pooled effects of acupuncture met the threshold for minimal clinically important differences, defined as the smallest differences in scores that people with OA would perceive to be beneficial.[62] The clinically relevant effects for knee osteoarthritis have been estimated to be standardized mean differences of 0.39 for WOMAC pain and 0.37 for WOMAC function.[62]

To further aid clinical interpretation, we also converted the pooled standardized mean differences to the percent changes from baseline of acupuncture relative to the comparison control, using the approach recommended by the Cochrane Musculoskeletal Group. For these calculations, we first calculated the absolute change by multiplying the pooled standardized mean difference by the standard deviation of the control group of the trial that has a large weighting in the overall result in RevMan, and is most representative, in terms of the patient characteristics and the baseline mean and standard deviation of the control group. This absolute change was then divided by the baseline value of this trial's control group to calculate a relative percent change from baseline. If there is "considerable or substantial heterogeneity",[60] it may be misleading to quote an

average clinical value for the intervention's effect, and in such cases, we did not compute pooled absolute and relative percent changes.

If any studies within any category reported insufficient data for pooling even after asking authors for more details, we excluded such studies from the meta-analyses and described their results narratively.

If at least six studies were available for a meta-analysis, we assessed the likelihood of small study bias by constructing funnel plots.[63]

Acupuncture adequacy assessments

Two acupuncturists (LL, Marcos Hsu), who have a combined acupuncture clinical experience of nearly forty years in treating knee OA, and who have both previously worked on RCTs and systematic reviews of acupuncture, independently assessed the adequacy of the acupuncture administered in the trials. Consensus was achieved by discussion. The adequacy of four aspects of the acupuncture[64] were assessed: choice of acupuncture points; number of sessions; needling technique; and acupuncturist's experience. The adequacy of the sham intervention was also assessed, using an open-ended question. The acupuncturists' assessments of adequacy were based only on the description of the study population and the acupuncture procedure; the assessors were blinded to the results of the study and the publication (author and journal).

The assessors had previously used this adequacy assessment instrument for the earlier systematic review,[36] of which this is an update. However, the acupuncturists decided that their previous assessments may have overestimated the adequacy of the acupuncture in some of the trials, and therefore, for this update, the acupuncturists assessed again the 11 trials previously assessed for the earlier version and also assessed the five new trials. For this update, the acupuncturists decided that the trials needed to include both an adequate number of treatments and also an adequate overall duration of treatment to earn a score of adequate on the 'number of sessions' item, whereas in the earlier review, only the number of treatments was considered, without regard for whether the duration of treatment (i.e., number of weeks of treatment) was of an adequate length. We asked the acupuncturists to guess the identity of each study being assessed to test the success of the blinding. The results of the tests of blinding to the results of the studies were based on the guesses of the studies made during the second adequacy assessment.

Subgroup and sensitivity analyses

Of the 11 elements of the Cochrane Back Review Group scale we conducted subgroup analyses only on the elements that relate specifically to randomization, blinding, and follow-up because we believed that these elements would be most likely to affect the effect of acupuncture for osteoarthritis. These six elements were the following: generation of random sequence; allocation concealment; baseline comparability; blinding success; acceptability of drop-out rate; and intention-to-treat analysis. For these subgroup tests, which were conducted among the sham-controlled trials only, which by design were all intended to have participants blinded, we used blinding success rather than patient blinding as the variable for the testing. We conducted another subgroup analysis on whether or not the sham used in the trial was likely to have analgesic/physiological activity, according to the acupuncturists' assessments. We also conducted a subgroup analysis on high ($\geq$ 6 out of 11) versus low quality trials, according to the Cochrane Back Review Group scale. Additionally, we conducted subgroup tests on the following seven clinical variables: the four variables related to acupuncture adequacy (see section 'Acupuncture adequacy

assessments' above), one variable related to the method of selecting acupuncture points (i.e., fixed formula or flexible formula), one variable related to electrical stimulation of acupuncture needles (Yes/No), and one variable related to length of follow-up (i.e., greater than or equal to versus less than three months from randomization).

For these subgroup analyses, we used a significance test, as described by Deeks et al[60,65] to investigate whether differences in effects of acupuncture between any two subgroups for any variable were statistically significant for either the pain or function outcome. For these subgroup analyses, we subgrouped all peripheral joint OA sham-controlled trials. Among the 15 potential variables for subgroup analyses described above, we conducted a statistical test for interaction only if there were more than one study in each subgroup, as is required.[65] For all subgroup comparisons, the P value for the interaction test was reported, as were the pooled estimates and I^2 values from each of the two relevant subgroups.[60]

We conducted one sensitivity analysis. That is, for the acupuncture versus sham comparison, for which the effect sizes were small and therefore potentially sensitive to use of different analytic methods, we calculated the standardized mean differences using both changes from baseline and post-treatment values. Namely, we used a comparison of means and standard deviations (SDs) of changes from baseline for each group for the primary analysis and we used a comparison of means and SDs of post-treatment scores for each group for a sensitivity analysis.[60] We conducted separate analyses using both changes from baseline and post treatment scores in order to test whether the statistical significance of the pooled standardized mean differences differed depending on the sets of values used.

RESULTS

Description of studies

Included studies

General study characteristics: Numbers of studies and participants; languages of publication; peripheral joints included; unpublished information obtained from authors:

We included 16 RCTs[1-3,43-48,54-59,66] representing 3498 randomized OA participants. Twelve of the RCTs included only people with OA of the knee, three included only people with OA of the hip,[44,46,47] and one included a mix of people with OA of the hip and/or OA of the knee.[59] All studies but one[43] were published in English. We obtained unpublished data from nine authors, including both outcome data[1,3,45,48,54,58] and methodological information.[3,45,55,56,66]

Characteristics of participants and details on outcome measures

All RCTs included primarily older participants, with a mean age of 60 or greater, and a mean duration of osteoarthritis knee pain of five years or more. For all RCTs, participants needed to be diagnosed with OA to be eligible, and all but two RCTs[2,45] required radiological evidence of OA. For all but two RCTs,[44,47] the WOMAC instrument was used to measure outcomes. (One other RCT[46] used a modified version of the WOMAC.) No RCTs reported that the OA diagnosis was made according to the principles of traditional Chinese medicine.

Method of selecting acupuncture points and stimulating needles

Five RCTs[45,47,48,54,58] used a flexible formula for point selection, and ten RCTs[1-3,43,44,46,55-57,66] used a set formula. For the one remaining pragmatic trial,[59] the point selection and needling technique were entirely at the discretion of the treating physicians. Superficial needling alone was used in one

trial,[2] whereas 13 trials[1,3,43-45,47,48,54-58,66] used sufficiently deep needle stimulation to elicit the *de qi* needling sensation. (Two trials did not report on *de qi*: the pragmatic trial[59] and one other small trial.[46]) Electrical stimulation of the needles was used in five trials,[1-3,47,55] and for three of these trials,[2,3,47] all needles were electrically stimulated.

Risk of bias in included studies

Among the ten RCTs that included a sham control (counting Sangdee 2002 * twice because it included two different comparisons for our meta-analysis, as described below, in the sub-section 'Results: effects of interventions: acupuncture versus sham'), we considered the five[1,3,45,54,58] with the highest quality ratings on the Cochrane Back Review Group scale[67] to comprise the bulk of the evidence for this review. These five RCTs have been published in leading international medical journals, all since 2004. Four of the five also included a waiting list or other active treatment control group.[1,45,54,58] Only two of the five had any obvious methodological flaws, which in both cases were due to higher dropout rates, in the sham group for one[3] and in the education control group for another.[1] Four of these trials[1,45,54,58] had a six month outcome assessment, but for only one[1] was a treatment schedule maintained up until the final six month measurement point.

For all sham-controlled RCTs, the schedule for the sham acupuncture procedure was the same as that for the true acupuncture procedure. We could not be certain that the shams used in three of the sham-controlled trials were sufficiently credible in fully blinding participants to the treatment being evaluated.[1-3] For all waiting list-controlled trials, participants on the waiting lists were allowed to receive the current level of oral NSAID or analgesic therapy.

Assessments of acupuncture adequacy

All of the trials included in this review were judged adequate on "Choice of acupoints" and "Needling technique", but only two of the trials[1,55] were judged adequate in terms of the acupuncturist's experience. For five of the trials,[45,46,48,56,66] the number of acupuncture sessions was judged inadequate. Also, for six trials,[43-45,54,58,66] the acupuncture adequacy assessors noted that the sham needling may have had physiologic activity.

Funding sources

Five RCTs did not report funding sources.[43,46,56,57,66] Five[1,3,45,47,55] were funded by government grants, 1[2] by a university, 1[48] by a hospital, 1[44] by a pharmaceutical company, and 3[54,58,59] by social health insurance companies.

Effects of interventions

Of the sixteen RCTs that met the selection criteria, all except three[43,47,56] reported extractable outcome data. Three trials included people with only hip OA,[44,46,47] but we did not meta-analyze these three trials together in a hip-specific analysis because there was heterogeneity of controls and outcome measures, and the outcomes were poorly reported or nonstandard. For knee OA, eleven trials reported extractable outcome data.

Acupuncture versus sham

Ten trials included a sham control, nine in people with knee OA and one in people with hip OA. Data for all but one[43] sham-controlled knee OA trial could be used for our quantitative analysis.

One sham-controlled trial[2] randomized participants to the following four groups: 'placebo tablet plus sham electroacupuncture', 'diclofenac tablet plus sham electroacupuncture', 'placebo tablet plus true electroacupuncture', and 'diclofenac tablet plus true electroacupuncture'. Because our meta-analysis addressed the question of whether acupuncture is more effective than sham, we included in our meta-analysis the following two comparisons from this trial as if they were from different studies, as recommended[68]: 'acupuncture versus sham with placebo tablet co-intervention'[2] and 'acupuncture versus sham with diclofenac tablet co-intervention'.[2] With this approach, intervention groups from this trial were only entered once in the meta-analysis.

In comparison with a sham control at the short-term follow-up, acupuncture showed improvements in OA pain (standardized mean difference -0.28, 95% confidence interval -0.45 to -0.11; 9 trials; 1835 participants; $I^2 = 64\%$) (Figure 1), function (-0.28, -0.46 to -0.09; 9 trials; 1829 participants; $I^2 = 69\%$) and symptom severity (-0.29, -0.50 to -0.09; 9 trials; 1767 participants; $I^2 = 74\%$), but the results were heterogeneous. The range of SMDs for the pain outcome was from -0.99 in the trial showing the greatest benefit to +0.05 in the trial showing no benefit (Figure 1). This corresponds to an absolute and relative percent improvement relative to a sham control of -29.06% and -48.03% in the trial showing the greatest benefit of acupuncture[3] to 0.90% and 2.02% in the trial showing no benefit of acupuncture.[45] Two out of the nine trials showed effect estimates higher than the pre-defined clinical relevance thresholds of 0.37 and 0.39 for pain and function respectively, but the pooled estimates were lower than the threshold. Results of the short-term meta-analyses of acupuncture versus sham for peripheral joint OA described above were unchanged when we restricted to only the knee-only trials (i.e., when we removed the Fink 2001 study, the one relevant hip osteoarthritis sham-controlled trial with extractable short-term data). In comparison with a sham acupuncture control at the six-month follow-up, acupuncture showed borderline statistically significant, clinically irrelevant improvements in knee OA pain (-0.10, -0.21 to 0.01; 4 trials; 1399 participants, $I^2 = 0\%$) (Figure 2), function (-0.11, -0.22 to 0.00, 4 trials; 1398 participants; $I^2 = 6\%$) and symptom severity (-0.11, -0.22 to 0.00, 4 trials; 1398 participants; $I^2 = 2\%$), and there was low heterogeneity.

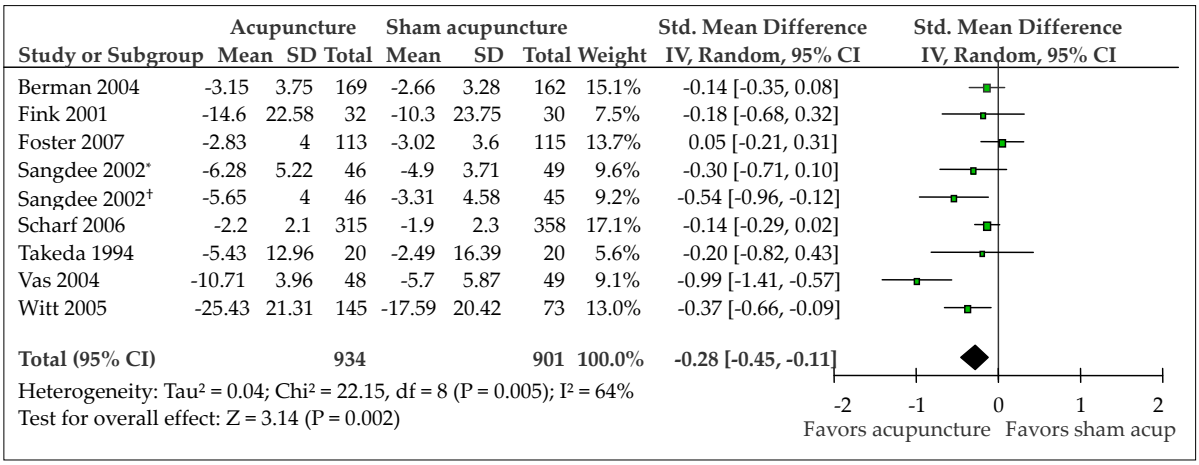


Figure 1. Effects of acupuncture versus a sham control group on the pain outcome at the short-term measurement point

*Comparison of electroacupuncture with sham acupuncture using a diclofenac co-intervention. †Comparison of electroacupuncture with sham acupuncture using a placebo diclofenac co-intervention.

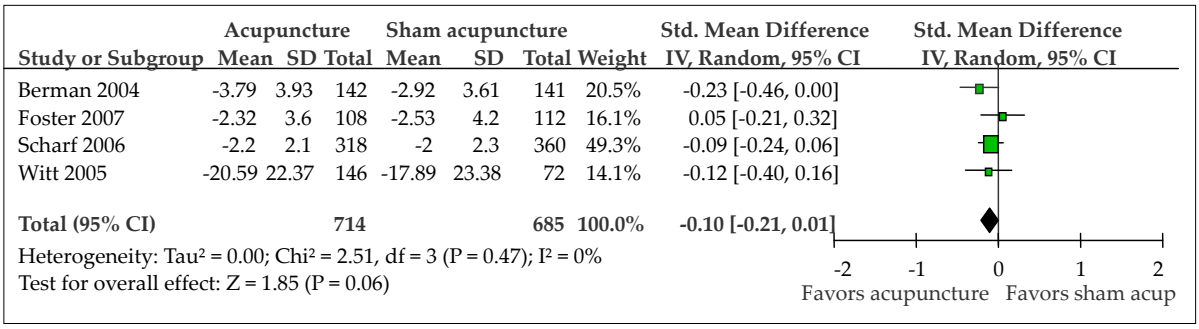


Figure 2. Effects of acupuncture versus a sham acupuncture control group on the pain outcome at the long term measurement point

Acupuncture versus waiting list

In comparison with a waiting list control, acupuncture was associated with clinically relevant short-term improvements in OA pain (-0.96, -1.19 to -0.72; 4 trials; 884 participants; $I^2 = 41\%$) (Figure 3), function (-0.89, -1.18 to -0.60; 3 trials; 864 participants; $I^2 = 64\%$), and symptom severity (-0.92,-1.16 to -0.67; 3 trials; 864 participants; $I^2 = 52\%$). The pooled effect estimate for the pain outcome corresponds to an absolute and relative percent improvement of -14.54% and -29.14% relative to a waiting list. There was moderate heterogeneity[60] but the benefits of acupuncture in each individual trial, as well as the pooled benefits, were much larger than our predefined thresholds for clinical relevance.

The effects of acupuncture were heterogeneous across trials that involved 'head-on' comparisons of acupuncture versus different active comparator controls (Figure 3). These head-on comparisons each included only a single trial so no meta-analysis could be performed. In the head-on comparisons of acupuncture with the 'supervised osteoarthritis education' control[1] and the 'physician consultations (with a physiotherapy co-intervention)' control,[54] acupuncture was associated with short- and long-term improvements in pain and function. In both cases, the benefits of acupuncture exceeded our thresholds for clinical relevance. In the head-on comparisons of acupuncture with the 'home exercises/advice leaflet alone',[48] and 'supervised exercise alone' [48] controls, there was also evidence that acupuncture was associated with similar treatment effects as the controls.

Acupuncture plus another active treatment versus other active treatment alone

The Foster 2007 trial included three treatment arms: an exercise based physiotherapy program (including supervised plus home exercises), exercise based physiotherapy program plus true acupuncture, and exercise based physiotherapy program plus non-penetrating acupuncture. For this trial,[45] which was the only trial with extractable outcome data that evaluated acupuncture as strictly an adjuvant to existing care, acupuncture as an adjuvant to an exercise based physiotherapy program (including supervised plus home exercises) did not result in any greater improvements than the exercise based physiotherapy program alone (Figure 3).

Trials not included in meta-analyses

Three RCTs,[43,47,56] one each using a sham, waiting list, and active treatment control, did not have extractable outcome data either because their results were reported in a way that could not be entered in the meta-analyses or because their results were difficult to interpret.

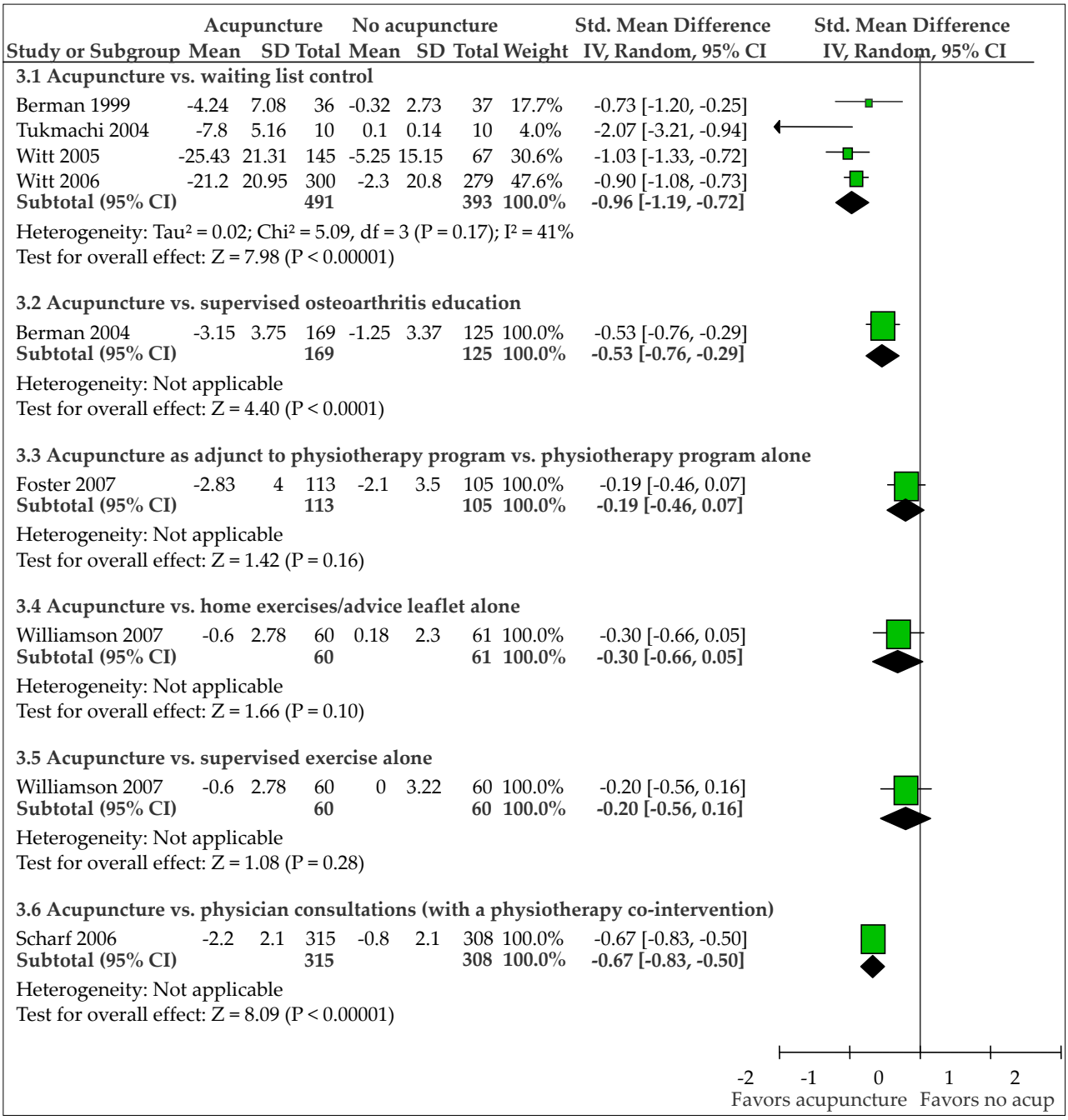


Figure 3. Effects of acupuncture compared with a waiting list or other active treatment control group on the pain outcome at the short-term measurement point

Safety of acupuncture

Eight RCTs[1,2,44-46,48,54,58] described adverse events across groups, and they found that the frequency of adverse events was similar between the acupuncture and control groups. Pooling of adverse events across these RCTs was not possible because of limited reporting and heterogeneous methods. No serious adverse events were reported to be associated with acupuncture. The frequency of minor side effects of acupuncture, primarily minor bruising and bleeding at needle insertion sites, ranged from 0%[44,45,55] to 45%.[2] These frequencies varied widely because of heterogeneous and scanty reporting and different definitions of what constitutes a side effect of

acupuncture versus what is an inherent part of treatment (for example, occasional bruising at needle insertion site).

Subgroup analyses

We planned subgroup analyses for 15 variables, but only eight had sufficient data available for a statistical test for interaction. Of the eight subgroup analyses on clinical and methodological variables, there was statistically significant heterogeneity of effect estimates between the two substrata for the following four variables for the pain outcome: success of blinding (Yes/Not sure); likely physiological activity of sham control (Yes/No); use of electrical stimulation of needles (Yes/No); and adequate number of acupuncture sessions (Yes/No) (see Table directly below). The function outcome subgroup analyses results were similar. The six trials that were judged to have been successfully blinded were the same six trials for which the acupuncture adequacy assessors noted that the sham needling may have had physiologic effects. When restricting to these successfully blinded trials with potentially physiologically active shams, the pooled results were smaller and only borderline statistically significant for the pain outcome and no longer statistically significant for the function outcome (-0.11, -0.29 to 0.07). Also, dividing trials into the two subgroups defined by this variable of blinding/sham type substantially reduced the overall heterogeneity (Table), suggesting that the type of sham used and the likelihood of whether or not it successfully blinds participants to treatment assignment may explain some of the overall heterogeneity of the sham controlled trials. For the 'sufficient number of sessions delivered over an adequate treatment duration (Yes/No)' and 'electrical stimulation (Yes/No)' criteria, the pooled estimates were statistically significant only in the subgroups that met either of the criteria (see Table below), but neither of these variables explained the heterogeneity.

Table. Results of the subgroup meta-analyses for the pain outcome*

	Variable	No. RCTs	No. pts	Effect size (95% CI)	I^2,%	Inter-action P†	Not met‡
	All trials	9	1835	-0.28 (-0.45,-0.11)	63.9	-	-
Method-ological variables	**Generation of random sequence**					0.141	SS
	Adequate	7	1649	-0.25 (-0.45,-0.05)	69		
	Unclear or no	2	186	-0.42 (-0.71,-0.13)	0		
	Allocation concealment					0.215	SSFi
	Adequate	6	1587	-0.26 (-0.48,-0.04)	74.2		
	Unclear	3	248	-0.36 (-0.61,-0.11)	0		
	Blinding success					**0.042**	BSSV
	Yes	5	1221	-0.15 (-0.28,-0.01)	14.5		
	Uncertain	4	614	-0.47 (-0.84,-0.10)	77.5		
	Intention-to-treat analysis					0.532	FiFoSST
	Yes	4	1319	-0.35 (-0.63,-0.07)	80.7		
	Unclear or no	5	516	-0.20 (-0.43,0.03)	35.2		

	Variable	No. RCTs	No. pts	Effect size (95% CI)	I^2,%	Inter-action P†	Not met‡
Clinical variables	**Sufficient no. of sessions delivered over adequate treatment duration**					**0.047**	FoT
	Yes	7	1567	-0.34(-0.54, -0.15)	66.1		
	No	2	268	0.01 (-0.23, 0.25)	0		
	Electrical stimulation was used with the acupuncture					**0.042**	BSSV met
	Yes	4	614	-0.50 (-0.81,-0.20)	66		
	No	5	1215	-0.11 (-0.29,0.07)	42.3		
	Physiological activity of sham, as judged by acupuncturist					**0.042**	BSSV
	Likely	5	1221	-0.15 (-0.28,-0.01)	14.5		
	Not likely	4	614	-0.47 (-0.84,-0.10)	77.5		
	Formula versus flexible formula for point selection					0.057	BFiSSTV used formula
	Formula	6	716	-0.39 (-0.66,-0.12)	72.8		
	Flexible formula	3	1119	-0.14 (-0.34,0.05)	77		

* Subgroup analyses for the sham controlled peripheral joint OA trials, based on changes of pain scores from baseline at the short-term (i.e., closest to 8 wks) follow up.

†A statistical test for interaction could not be performed for the following prespecified subgroups because there were one or fewer studies in one of the strata of each of these subgroups: comparability of baseline, acceptability of drop-out rate, CBRG score $\geq$ 6, follow-up length $\geq$ 3 months, choice of acupoints, needling technique, acupuncturists' experience.

‡"Not met" column lists the first (or first two) letters of the author of the studies that do not meet the criterion.

Sensitivity analyses

Sensitivity analyses showed that the pooled standardized mean differences calculated using differences in post-treatment scores between groups were slightly larger than the pooled standardized mean differences calculated using differences in improvements between groups, for the acupuncture versus sham comparison; this was attributed to the slightly better baseline WOMAC scores in the acupuncture groups than the sham groups, for three RCTs.[3,54,58] However, there were no outcomes at any time points for which the statistical significance of the pooled result changed depending on whether post-treatment scores or differences in improvements were used. The pooled standardized mean differences calculated using the fixed-effect analysis were slightly smaller than the pooled standardized mean differences using the random effects analysis, for the acupuncture versus sham short-term comparison. However, the statistical significance of the pooled result did not change depending on whether the random-effects or fixed-effect analyses were used.

The only reported adverse events attributable to the acupuncture were slight bruising or bleeding at acupuncture points.

DISCUSSION

Summary of main results

Sham-controlled trials are designed to minimize placebo effects, and thereby measure the true biological effects of acupuncture. Currently available sham-controlled trials show statistically significant pooled benefits of acupuncture relative to sham; however, these benefits are small and do not meet our pre-defined criteria for clinical relevance. Also, some sham-controlled trials showed no benefit of acupuncture over sham and there was substantial heterogeneity of results, which may be due to differences in sham interventions, differences in acupuncture protocols, differences in settings, and varying proficiencies of the acupuncturists. One trial that evaluated acupuncture as an adjuvant to a standardized exercise/advice program found no additional benefit of acupuncture.[45] Head to head trials of acupuncture versus other active treatments have had heterogeneous results: two trials showed no statistically significant differences between acupuncture and exercise intervention controls (Williamson 2007 - two different exercise controls used), while two other trials showed a statistically significant benefit of acupuncture compared to the active intervention controls of 'supervised osteoarthritis education'[1] and 'physician consultation (with a physiotherapy co-intervention)'.[54] However, these two latter trials are difficult to interpret because both the acupuncture and the sham arms showed benefits over the active treatment controls, suggesting placebo effects may have played a role. Currently available trials of acupuncture for osteoarthritis suggest statistically significant and clinically relevant benefits of acupuncture compared to a waiting list control. We considered patient blinding to be the most critical factor related to the applicability of the evidence and the risk of bias of the included studies. Therefore, the remainder of the Discussion focuses primarily on the design of the included trials, and particularly the different types of controls used, and how this variability of controls may explain the variability in the findings.

Limitations and challenges of the sham control design

Using a sham acupuncture comparison is intended to control for patient expectations and placebo effects, and thereby estimate the effects of acupuncture due solely to the point specific placement of the acupuncture needles. However, there are two considerations that must be borne in mind in designing a sham control. First, it is important to design a sham that is physiologically inert. Yet, there are no standards by which to determine which 'sham' point locations, depths or directions of sham needle insertion, or durations of sham needle placement will truly result in no physiological activity, and which are therefore truly inert. However, many authorities agree that in order to be inert, sham needles should not penetrate the skin.[69] The reason that this is important is because one commonly proposed mechanism suggests that acupuncture (as well as sham acupuncture), may work through a diffuse noxious inhibitory control (DNIC) mechanism. That is, the DNIC theory suggests that noxious stimuli (e.g., needles) applied to any part of the body can produce analgesic effects, even at distant sites.[70,71] Because inserting needles, even at non-acupoints, may produce an analgesic effect, needle insertive sham acupuncture may not be an appropriate control, and may be better understood as a "poor form of acupuncture treatment".[71] Indeed, physiological activity of needle penetration, even if superficial and at non-acupuncture points, is suggested by several lines of research,[69] including RCTs showing larger effects of a superficially-penetrating needle acupuncture than a non-penetrating sham control[2,72]; positron emission tomography research indicating sham acupuncture can stimulate regions of the brain associated with natural opiate production[73]; and animal studies showing sham needle-insertion can have non-specific analgesic

effects through the postulated DNIC mechanism.[70] One possible explanation for our subgroup analysis which showed smaller benefits of acupuncture when compared with a sham involving needle penetration (versus a sham with no needle penetration) is that needle penetrating shams may have physiological activity, even if the needles are inserted only superficially. Indeed, superficial needling is a common technique used in many authentic traditional Japanese acupuncture styles.[74]

The second important consideration in designing an appropriate sham is somewhat in tension with the first: the sham must be sufficiently believable to participants as an authentic treatment to fully allow for the control of placebo effects. Developing a sham that is believable to trial participants and yet that does not involve needles that penetrate the skin is a challenge. One way to increase participants' beliefs in the authenticity of the sham may be to inform participants of the sham and true acupuncture treatment arms in a manner that suggests two different real, active treatments are compared, without mentioning terms such as placebo or sham.[58] However, this disclosure practice may not be permitted by some internal review boards, and some bioethicists[75] have argued that this practice violates the ethical requirements of informed consent.

To test whether a sham was believable to the participants, it has been proposed that investigators should ask participants to guess their treatment assignment and/or guess whether or not they received an authentic treatment or a sham.[76] However, the validity of testing credibility of the sham is unclear and deserves further study because correctly guessing treatment assignment could be highly correlated with a treatment's effectiveness, or lack thereof.[77-80] That is, if only true acupuncture and not sham has a true physiological benefit, then one might expect that participants assigned to true acupuncture would be more likely to guess that they were receiving an authentic treatment than would participants assigned to the sham acupuncture because only the true acupuncture participants would be experiencing the beneficial effects of the treatment. Or as Altman and colleagues describe it, "end of trial tests of blindness might be tests of hunches for adverse effects or efficacy".[77] Indeed, both Altman and colleagues[77] as well as David Sackett[79,81] have argued that trial investigators should vigorously test for blinding before trials start, but that tests of blinding at the end of trials can be difficult to interpret, and therefore cannot be recommended in all circumstances. Nevertheless, in this review, we have considered the results of blinding tests after the trial has begun because acupuncture is a particularly difficult intervention for which to develop a believable sham, and it is possible that the trial participants would continue to try to figure out whether they were receiving the true or sham treatment while the trial is underway. While we have considered the results of such post-treatment blinding tests, we recognize that these results can be uninterpretable.

Possible explanations for the findings of the sham-controlled trials

For the two sham-controlled RCTs that found clinically relevant benefits of acupuncture,[2,3] the credibility of the sham was not tested and the informed consent procedure was not described. In one of these RCTs[2] the sham involved patch electrodes attached to the same four knee points used in the true acupuncture group, with mock electrical stimulation. Because no needles were used, participants randomized to sham probably did not believe they were receiving traditional acupuncture. Nevertheless, these participants may have believed they were receiving an authentic treatment because the non-needle sham used in this trial is comparable to a sham that was shown to be similarly credible to authentic acupuncture in a previous RCT.[82] In the other sham-controlled RCT with highly positive results,[3] some unblinding may have been possible: the sham needles did

not penetrate the skin and therefore may have been less believable to participants than a sham that involves needle penetration of the skin. While a non-penetrating sham needle similar to the one used in this trial[3] has been shown to be indistinguishable from the real penetration of a needle among acupuncture-naïve participants in an earlier validation study,[83] the credibility of the specific sham needle used in this trial[3] was not tested among the participants included (who were also acupuncture naïve) so we cannot be certain that all the participants in this trial believed this sham to be an authentic treatment. Although some unblinding may have been possible in these two sham-controlled RCTs, a possible alternative explanation for the fact that these were the only two sham-controlled trials that clearly showed clinically relevant benefits is that only these two sham-controlled trials used intensive electrical simulation at all local knee points; electrical stimulation of needles is not always used because of logistical and cost constraints, but it may produce stronger analgesic effects than manual stimulation of needles.[84]

A third large trial[1] found small, statistically significant benefits of acupuncture relative to sham but these benefits did not meet our threshold for clinical relevance. This trial used an innovative sham that involved a combination of penetrating and non-penetrating needles. Most participants believed they received true acupuncture at the four week credibility test, but at the 26 week test, the sham group participants were more likely than the acupuncture group participants to guess that they received a sham. At the 26 week test, this may be partly explained by the fact that the participants receiving sham were experiencing no benefit, as further described above.

Two other large sham-controlled RCTs[54,58] used fully needle-penetrating shams, which involved an average of 10[54] to 13[58] needles, inserted superficially at non-acupuncture points. These two RCTs[54,58] found clinically irrelevant[54] or minimally clinically relevant[58] differences between acupuncture and sham, but clinically relevant differences between both acupuncture groups (i.e., both true and sham acupuncture) and the non-acupuncture comparison group. While the fully needle-penetrating shams used in these two RCTs were sufficiently similar to true acupuncture to ensure adequate blinding, at the same time, these shams were so similar to true acupuncture that they may have had weak physiological activity and not been true placebos.[69] Indeed, in another recent systematic review on acupuncture for knee osteoarthritis,[37] the reviewers judged that the shams used in these two RCTs were probably physiologically active and inappropriate as controls.

In these two RCTs,[54,58] the trial participants may have had a positive attitude towards acupuncture and high expectations of a benefit,[85] but less enthusiasm for the control treatments, which are standard care therapies that the trial participants could have easily obtained on their own. Also, these two RCTs were funded by German insurance agencies to determine whether acupuncture should be reimbursable.[86] Participating people with OA and their physicians probably knew that patients would not need to pay for their acupuncture treatments out of pocket in the future if these RCTs had positive results.[87] As a result, participants may have overstated benefits of the treatment they believed to be acupuncture, which in these well-blinded trials[54,58] were both the true and sham acupuncture treatments, and understated assessments of non-acupuncture controls. Any such biased assessments may have contributed to the large differences between the acupuncture (i.e., both true or sham) and the non-acupuncture control and the small differences between the acupuncture and sham groups.

The most recent sham-controlled trial[45] used a sham that involved non-penetrating acupuncture at the same points as the true acupuncture, with both true and sham acupuncture administered over a duration of three weeks. In this methodologically sound trial, the sham was found to be highly credible and believable as an authentic treatment. However, because the non-

insertive sham needles were placed at the true acupuncture points, these sham needles may have had some physiological activity due to a massaging effect on the acupuncture points. Indeed, the investigators found that a considerable proportion of participants in the sham group experienced the *de qi* sensation, and they noted that "we cannot consider this [sham] intervention as inert."

As noted above, for only two trials[2,3] did the acupuncture meet the thresholds for clinical relevance relative to the sham; however, these thresholds should be interpreted with caution because the sham comparator may be an active treatment in itself rather than an inert placebo. Also, these clinical relevance thresholds, while helpful as general guidelines for estimating whether OA treatment effects are meaningful, cannot be considered fixed across all different clinical situations, and may vary depending on the characteristics of the population, the condition studied, the types of interventions, and the types of controls. Furthermore, data on clinical relevance of benefits need to be considered together with data on costs and adverse reactions. Clinicians must weigh all of this information to decide whether estimates of benefit are important for their patient population, and in their setting.

Limitations of the waiting list and other active treatment control designs

The major limitation of the waiting list and the other active treatment control designs is the lack of blinding. The clinically relevant benefits of acupuncture in comparison to the waiting list and, in some cases, the other active treatment controls, might be partly attributable to either non-specific effects associated with the patient-acupuncturist relationship[88,89] (e.g., attention, compassionate care[90]) or to expectations of a benefit by participants.[91] That is, study participants know they are getting acupuncture and might expect acupuncture to benefit them more than other active therapies more commonly used, perhaps because of the novelty of acupuncture, its ritualistic associations, or its ancient history of use.[88,92] Therefore study participants receiving acupuncture may report feeling better, regardless of whether the acupuncture worked. In fact, the placebo effects of acupuncture are expected to be even stronger than the placebo effects of conventional active treatments,[88,92,93] which is a limitation of head-on comparisons. However, despite the limitations of the head-on and waiting list comparator designs, these designs may still best approximate the average likely response to acupuncture in clinical practice, in which treatment effects and placebo factors, expectation effects, and patient preferences may all operate.[94]

Possible interpretations of subgroup analyses findings

Our subgroup analyses on blinding success suggests that the benefits of acupuncture relative to sham are smaller and borderline to non-significant in the sham-controlled trials that are judged most likely to adequately blind participants to treatment assignment. However, this finding can be interpreted in at least two different ways. The first interpretation is that acupuncture is mostly a placebo, and that when acupuncture is compared with a credible sham that adequately controls for the placebo effect, there is little if any remaining benefit of acupuncture. However, a second interpretation is that the effects of acupuncture are underestimated when compared with such believable shams because such truly believable shams are often not inert placebos, but have physiological effects of their own, due either to insertion of needles, even if superficially inserted at non-acupuncture points, or placement of needle-like sham devices at true acupoints.

Furthermore, for this subgroup analysis, we considered the success of blinding to be uncertain for the trials that either did not test blinding success[2,3] or that showed differential awareness of group membership among participants.[1] We might instead have assumed that all the shams used

in these trials were successful at blinding participants to treatment assignment. Indeed, few systematic reviews and RCTs even consider tests of blinding success, and instead it is typically assumed that placebo controlled trials are adequately blinded. For example, in a sample of 97 placebo-controlled RCTs from five leading general medical journals, only seven RCTs provided evidence on the success of blinding, and five of these seven reported that the success of blinding was imperfect.[76]

Our subgroup analyses also showed that trials judged not to have an adequate number of treatment sessions delivered over a sufficient duration showed smaller benefits compared with trials with an adequate number of sessions. This finding agrees with that of an earlier meta-analysis,[95] which found that fewer than six acupuncture treatments was significantly associated with smaller benefits, even after adjusting for study quality.[95] Finally, our subgroup analyses found that trials using electrical simulation of needles showed a larger benefit than trials using only manual stimulation. This finding agrees with mechanistic studies of acupuncture, which have similarly suggested that electrical stimulation produces stronger analgesic effects than manual stimulation alone.[84]

All of our subgroup analyses should be considered hypothesis generating only and not confirmatory, both because of the small number of trials relative to the large number of subgroup comparisons tested, and also because of the possibility of confounding. For example, the subgroup finding of a smaller benefit in the trials judged most likely to adequately blind participants might be confounded by the fact that the trials that tested success of blinding were the same trials that were conducted at multiple centers and that involved a large number of treating physicians or physiotherapists trained in acupuncture. A potential limitation of such large scale multi-center acupuncture trials is that, with an increasing number of study sites and practitioners, the capacity to implement and monitor adherence to the acupuncture and sham protocols could be reduced, potentially increasing the risk of a Type II error.

AUTHORS' CONCLUSIONS

Implications for practice

The effects of true acupuncture relative to sham do not meet our pre-specified thresholds for clinical relevance.[62] Thus, the effects of acupuncture relative to sham acupuncture are too small to be perceived by participants as beneficial[62]; however, few if any other commonly used treatments for osteoarthritis meet these thresholds[62] for minimal clinically important differences.[96] For example, NSAIDS (relative to an inert placebo) do not meet these thresholds,[10,62] yet NSAIDs are used regularly by half of all people with painful osteoarthritis.[10] Acupuncture, in contrast, is used by only about 1%[16] of people with osteoarthritis, and most of these people do not use it specifically for treating their osteoarthritis.[16]

The effects of true acupuncture relative to a waiting list control and some of the other active treatment control groups do exceed our thresholds for clinical relevance.[62] The only other non-pharmacological treatment for osteoarthritis with benefits close to or exceeding the thresholds for clinical relevance is exercise, with standardized mean differences of .39 for pain and .31 for function, relative to a non-exercise control group.[97] However, in both cases, sham treatments were not used as controls, so some of the benefits measured may be attributable to expectation or placebo effects. Although exercise cannot be compared with sham exercise, acupuncture can be compared with sham acupuncture, although sham acupuncture may not be an inert placebo. While the comparison of acupuncture with sham, which shows very small benefits of acupuncture at

best, is useful for estimating the specific biological effects of acupuncture, it may be less relevant for clinical applications. Rather, the evaluation of the whole package of acupuncture, including both its specific and non-specific components (as is the case with exercise and education), may be of equal or greater clinical relevance. Overall, the studies suggest that people with osteoarthritis find meaningful benefits through acupuncture, although these benefits may be largely mediated through placebo effects.

The fact that few if any OA treatments have specific effects that meet the threshold for clinically relevant benefits[96] should not be interpreted to mean that we simply have no effective treatment for osteoarthritis. Rather, it may be that the threshold for clinical relevance is too high for any individual treatment alone, and that a multidisciplinary approach to OA patient management, with a focus on combining several nonpharmacological therapies[35] is necessary. Some clinicians and patients may consider acupuncture as one treatment option in such a multidisciplinary integrative approach[35] to treating knee osteoarthritis.

The relative benefits of acupuncture compared with other treatments cannot be reliably assessed because there is a scarcity of direct comparisons. Comparing different OA treatments by using indirect comparisons of effect sizes from different meta-analyses can be misleading because of differences in the numbers of studies, comparators used, and characteristics of participants.[12,98] Indeed, the recent Osteoarthritis Research International recommendations state that at best we can only examine whether there is no overlap of the 95% confidence intervals between the meta-analytic effect sizes of different treatments to see whether there may be differential benefits.[12] And yet because for most OA treatments there are small effect sizes with wide confidence intervals, and differences in point estimates across different meta-analyses evaluating the same treatments, it is unreliable to estimate the relative effects of acupuncture compared to other active treatments using indirect comparisons.

Safety and costs are other considerations. Safety is best determined with large prospective surveys of practitioners, and three such surveys[31-33] show that serious adverse events after acupuncture are rare. There were no adverse events associated with acupuncture in this review, although heterogeneous reporting and relatively small sample sizes limit the usefulness of this safety data. In addition to efficacy and safety, people with OA and their clinicians will also need to consider costs because acupuncture treatment often needs to be paid for out of pocket, at least in part.

Implications for research

Considering the prevalence of knee osteoarthritis and its burden on the health system and society in general and the dearth of safe and effective treatments, it seems warranted to conduct additional RCTs evaluating the cost-effectiveness of acupuncture, as well as its short- and long-term effects relative to other active treatments and shams. Pragmatic comparisons (including cost-effectiveness studies) are now of particular clinical relevance, and some future trials should perhaps shift from sham controls to active controls. Also, future trials might shift focus from the knee to other peripheral joints, for which the current evidence is scarce.

The results of this systematic review may help inform the design of future trials in several ways. First, current RCT results[1,58] suggest that benefits may attenuate over time, and therefore for future trials that assess long-term outcomes, it may be important to maintain monthly acupuncture treatments in the months prior to the long-term assessment. Second, our sensitivity analysis suggests that electrical stimulation may be associated with better outcomes, and the two sham-

controlled trials in this review that used intensive electrical stimulation of all local knee points showed the greatest benefits.[2,3] While these findings might indicate a superiority of electroacupuncture over needle acupuncture without electrical stimulation, the finding may also be explained by the fact that electroacupuncture is probably more difficult to blind than needle acupuncture and some of the extra benefit seen with electroacupuncture may be due to incomplete blinding or placebo effects. Third, our sensitivity analysis suggests that an adequate number of treatments delivered over a time period of a sufficient duration may be associated with better outcomes. Fourth, acupuncture may elicit a greater placebo effect or meaning response[99] than usual care therapies, particularly among participants who have a preference for acupuncture, and therefore investigators conducting future pragmatic trials that compare acupuncture with other active therapies might consider asking participants about their preferences and expectations (before and after the intervention), and studying the potential effects of pre-treatment preferences on study outcomes. Furthermore, to minimize the recruitment of participants with a preference for acupuncture, advertisements to recruit participants should ideally not specify that acupuncture is one of the treatments being investigated. Fifth and last, our review suggests that skin-penetrating needle shams may be best at insuring blinding success, but that such penetrating shams may also have physiological activity. Future trials should therefore consider the use of non-insertive shams; however, because such non-insertive shams may be less believable to participants, if they are used, their credibility should be tested, certainly before the trial starts, and perhaps also during the trial.

There are at least three large and rigorous ongoing sham-controlled trials, all of which should be published within the next couple of years. The largest of these three trials, which was recently presented at a conference,[100] found no difference between true and sham acupuncture, but found significant differences between both the true and sham acupuncture groups and the waiting list control group.[100] The results of the two other sham-controlled trials currently ongoing will be unlikely to shift the currently very small pooled benefits of acupuncture relative to sham towards the threshold for clinical relevance; however, these ongoing trials, likely to be successfully blinded, will be important to further assess how much of the currently observed benefit of acupuncture relative to sham is due to expectation or placebo effects and how much is due to specific effects of the needle placement. However, the truth about acupuncture effects will always be difficult to assess, even through a systematic review of well-designed and well-reported RCTs. The complexities and potential biases inherent to both the non-acupuncture and sham acupuncture control designs makes it difficult to evaluate the subjective, patient-reported outcomes in peripheral joint osteoarthritis.

REFERENCES

1. Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. *Ann Intern Med* 2004;141(12):901-10.
2. Sangdee C, Teekachunhatean S, Sananpanich K, Sugandhavesa N, Chiewchantanakit S, Pojchamarnwiputh S, Jayasvasti S. Electroacupuncture versus diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. *BMC Complement Altern Med* 2002;2:3.
3. Vas J, Mendez C, Perea-Milla E, Vega E, Panadero MD, Leon JM, Borge MA, Gaspar O, Sanchez-Rodriguez F, Aguilar I, et al. Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. *BMJ* 2004;329:1216-20.
4. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, Kington RS, Lane NE, Nevitt MC, Zhang Y, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med* 2000;133(8):635-46.
5. Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, Heyse SP, Hirsch R, Hochberg MC, Hunder GG, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum* 1998;41(5):778-99.
6. Prevalence of disabilities and associated health conditions among adults--United States, 1999. *MMWR Morb Mortal Wkly Rep* 2001;50(7):120-5.
7. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Ann Rheum Dis* 2001;60(2):91-7.
8. Ausiello JC, Stafford RS. Trends in medication use for osteoarthritis treatment. *J Rheumatol* 2002;29(5):999-1005.
9. Wegman A, van der Windt D, van Tulder M, Stalman W, de Vries T. Nonsteroidal antiinflammatory drugs or acetaminophen for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines. *J Rheumatol* 2004;31(2):344-54.
10. Bjordal JM, Ljunggren AE, Klovning A, Slordal L. Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: meta-analysis of randomised placebo controlled trials. *BMJ* 2004;329(7478):1317.
11. Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. *Cochrane Database Syst Rev* 2006(1):CD004257.
12. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;16(2):137-62.
13. Larson AM, Polson J, Fontana RJ, Davern TJ, Lalani E, Hynan LS, Reisch JS, Schiodt FV, Ostapowicz G, Shakil AO, et al. Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology* 2005;42(6):1364-72.
14. National Collaborating Centre for Chronic Conditions. Osteoarthritis: national clinical guideline for care and management in adults. London: Royal College of Physicians, 2008.
15. Herman CJ, Allen P, Hunt WC, Prasad A, Brady TJ. Use of complementary therapies among primary care clinic patients with arthritis. *Prev Chronic Dis* 2004;1(4):A12.
16. Quandt SA, Chen H, Grzywacz JG, Bell RA, Lang W, Arcury TA. Use of complementary and alternative medicine by persons with arthritis: results of the National Health Interview Survey. *Arthritis Rheum* 2005;53(5):748-55.
17. Rao JK, Mihaliak K, Kroenke K, Bradley J, Tierney WM, Weinberger M. Use of complementary therapies for arthritis among patients of rheumatologists. *Ann Intern Med* 1999;131(6):409-16.
18. Towheed TE, Maxwell L, Anastassiades TP, Shea B, Houpt J, Robinson V, Hochberg MC, Wells G. Glucosamine therapy for treating osteoarthritis. *Cochrane Database Syst Rev* 2005(2):CD002946.
19. Cibere J, Kopec JA, Thorne A, Singer J, Canvin J, Robinson DB, Pope J, Hong P, Grant E, Esdaile JM. Randomized, double-blind, placebo-controlled glucosamine discontinuation trial in knee osteoarthritis. *Arthritis Rheum* 2004;51(5):738-45.
20. Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, Bradley JD, Bingham CO, 3rd, Weisman MH, Jackson CG, et al. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. *N Engl J Med* 2006;354(8):795-808.
21. McAlindon T, Formica M, LaValley M, Lehmer M, Kabbara K. Effectiveness of glucosamine for symptoms of knee osteoarthritis: results from an internet-based randomized double-blind controlled trial. *Am J Med* 2004;117(9):643-9.
22. Rozendaal RM, Koes BW, van Osch GJ, Uitterlinden EJ, Garling EH, Willemsen SP, Ginai AZ, Verhaar JA, Weinans H, Bierma-Zeinstra SM. Effect of glucosamine sulfate on hip osteoarthritis: a randomized trial. *Ann Intern Med* 2008;148(4):268-77.
23. Reichenbach S, Sterchi R, Scherer M, Trelle S, Burgi E, Burgi U, Dieppe PA, Juni P. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. *Ann Intern Med* 2007;146(8):580-90.
24. World Health Organization. A proposed standard international acupuncture nomenclature: report of a WHO scientific group. Geneva, 1991.

25. Lewith GT, Kenyon JN. Physiological and psychological explanations for the mechanism of acupuncture as a treatment for chronic pain. *Soc Sci Med* 1984;19(12):1367-78.
26. Han JS. Acupuncture: neuropeptide release produced by electrical stimulation of different frequencies. *Trends Neurosci* 2003;26(1):17-22.
27. Hui KK, Liu J, Marina O, Napadow V, Haselgrove C, Kwong KK, Kennedy DN, Makris N. The integrated response of the human cerebro-cerebellar and limbic systems to acupuncture stimulation at ST 36 as evidenced by fMRI. *Neuroimage* 2005;27(3):479-96.
28. Li A, Lao L, Wang Y, Xin J, Ren K, Berman BM, Tan M, Zhang R. Electroacupuncture activates corticotrophin-releasing hormone-containing neurons in the paraventricular nucleus of the hypothalammus to alleviate edema in a rat model of inflammation. *BMC Complement Altern Med* 2008;8:20.
29. Cherkin DC, Sherman KJ, Deyo RA, Shekelle PG. A review of the evidence for the effectiveness, safety, and cost of acupuncture, massage therapy, and spinal manipulation for back pain. *Ann Intern Med* 2003;138(11):898-906.
30. Lao L, Hamilton GR, Fu J, Berman BM. Is acupuncture safe? A systematic review of case reports. *Altern Ther Health Med* 2003;9(1):72-83.
31. MacPherson H, Thomas K, Walters S, Fitter M. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;323(7311):486-7.
32. Melchart D, Weidenhammer W, Streng A, Reitmayr S, Hoppe A, Ernst E, Linde K. Prospective investigation of adverse effects of acupuncture in 97 733 patients. *Arch Intern Med* 2004;164(1):104-5.
33. White A, Hayhoe S, Hart A, Ernst E. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. *BMJ* 2001;323(7311):485-6.
34. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. *Arthritis Rheum* 2000;43(9):1905-15.
35. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, Gunther K, Hauselmann H, Herrero-Beaumont G, Kaklamanis P, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis* 2003;62(12):1145-55.
36. Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for osteoarthritis of the knee. *Ann Intern Med* 2007;146(12):868-77.
37. White A, Foster NE, Cummings M, Barlas P. Acupuncture treatment for chronic knee pain: a systematic review. *Rheumatology (Oxford)* 2007;46(3):384-90.
38. Kwon YD, Pittler MH, Ernst E. Acupuncture for peripheral joint osteoarthritis: a systematic review and meta-analysis. *Rheumatology (Oxford)* 2006;45(11):1331-7.
39. Birch S. Systematic reviews of acupuncture -- are there problems with these? *Clinical acupuncture and oriental medicine* 2001;2:17-22.
40. Glanville JM, Lefebvre C, Miles JN, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. *J Med Libr Assoc* 2006;94(2):130-6.
41. Ernst E. Acupuncture as a symptomatic treatment of osteoarthritis. A systematic review. *Scand J Rheumatol* 1997;26(6):444-7.
42. Ezzo J, Hadhazy V, Birch S, Lao L, Kaplan G, Hochberg M, Berman B. Acupuncture for osteoarthritis of the knee: a systematic review. *Arthritis Rheum* 2001;44(4):819-25.
43. Molsberger A, Bowing G, Jensen KU, Lorek M. [Acupuncture treatment for the relief of gonarthrosis pain-a controlled clinical trial.] *Der Schmerz* 1994;8(1):37-42.
44. Fink MG, Wipperman B, Gehrke A. Non-specific effects of traditional Chinese acupuncture in osteoarthritis of the hip. *Complement Ther Med* 2001;9(2):82-9.
45. Foster NE, Thomas E, Barlas P, Hill JC, Young J, Mason E, Hay EM. Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: randomised controlled trial. *BMJ* 2007;335(7617):436.
46. Haslam R. A comparison of acupuncture with advice and exercises on the symptomatic treatment of osteoarthritis of the hip--a randomised controlled trial. *Acupunct Med* 2001;19(1):19-26.
47. Stener-Victorin E, Kruse-Smidje C, Jung K. Comparison between electro-acupuncture and hydrotherapy, both in combination with patient education and patient education alone, on the symptomatic treatment of osteoarthritis of the hip. *Clin J Pain* 2004;20(3):179-85.
48. Williamson L, Wyatt MR, Yein K, Melton JT. Severe knee osteoarthritis: a randomized controlled trial of acupuncture, physiotherapy (supervised exercise) and standard management for patients awaiting knee replacement. *Rheumatology (Oxford)* 2007;46(9):1445-9.
49. Angst F, Ewert T, Lehmann S, Aeschlimann A, Stucki G. The factor subdimensions of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) help to specify hip and knee osteoarthritis. a prospective evaluation and validation study. *J Rheumatol* 2005;32(7):1324-30.
50. Bellamy N. The WOMAC Knee and Hip Osteoarthritis Indices: development, validation, globalization and influence on the development of the AUSCAN Hand Osteoarthritis Indices. *Clin Exp Rheumatol* 2005;23(5 Suppl 39):S148-53.

51. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15(12):1833-40.
52. Escobar A, Quintana JM, Bilbao A, Azkarate J, Guenaga JI. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Western Ontario and McMaster Universities Osteoarthritis Index. *Clin Rheumatol* 2002;21(6):466-71.
53. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, Theiler R. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. *Z Rheumatol* 1996;55(1):40-9.
54. Scharf HP, Mansmann U, Streitberger K, Witte S, Kramer J, Maier C, Trampisch HJ, Victor N. Acupuncture and knee osteoarthritis: a three-armed randomized trial. *Ann Intern Med* 2006;145(1):12-20.
55. Berman BM, Singh BB, Lao L, Langenberg P, Li H, Hadhazy V, Bareta J, Hochberg M. A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. *Rheumatology (Oxford)* 1999;38(4):346-54.
56. Christensen BV, Iuhl IU, Vilbek H, Bulow HH, Dreijer NC, Rasmussen HF. Acupuncture treatment of severe knee osteoarthrosis. A long-term study. *Acta Anaesthesiol Scand* 1992;36(6):519-25.
57. Tukmachi E, Jubb R, Dempsey E, Jones P. The effect of acupuncture on the symptoms of knee osteoarthritis--an open randomised controlled study. *Acupunct Med* 2004;22(1):14-22.
58. Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, Hummelsberger J, Walther HU, Melchart D, Willich SN. Acupuncture in patients with osteoarthritis of the knee: a randomised trial. *Lancet* 2005;366(9480):136-43.
59. Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm. *Arthritis Rheum* 2006;54(11):3485-93.
60. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 (updated February 2008). The Cochrane Collaboration, 2008. Available from http://www.mrc-bsu.cam.ac.uk/cochrane/handbook500/ (accessed 2013 July 12).
61. Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336(7658):1413-5.
62. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol* 2002;29(1):131-8.
63. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;320(7249):1574-7.
64. Furlan AD, van Tulder MW, Cherkin DC, Tsukayama H, Lao L, Koes BW, Berman BM. Acupuncture and dry-needling for low back pain. *Cochrane Database Syst Rev* 2005(1):CD001351.
65. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors. Systematic Reviews in Health Care: Meta-Analysis in Context (2nd edition). London: BMJ Publication Group, 2001.
66. Takeda W, Wessel J. Acupuncture for the treatment of pain of osteoarthritic knees. *Arthritis Care Res* 1994;7(3):118-22.
67. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine* 2003;28(12):1290-9.
68. Higgins JPT, Deeks JJ, Altman DG (editors). Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 (updated February 2008). The Cochrane Collaboration, 2008. Available from http://www.mrc-bsu.cam.ac.uk/cochrane/handbook500/ (accessed 2013 July 12).
69. Birch S. A review and analysis of placebo treatments, placebo effects, and placebo controls in trials of medical procedures when sham is not inert. *J Altern Complement Med* 2006;12(3):303-10.
70. Le Bars D, Dickenson AH, Besson JM. Diffuse noxious inhibitory controls (DNIC). I. Effects on dorsal horn convergent neurones in the rat. *Pain* 1979;6(3):283-304.
71. Lewith GT, Machin D. On the evaluation of the clinical effects of acupuncture. *Pain* 1983;16(2):111-27.
72. Macdonald AJ, Macrae KD, Master BR, Rubin AP. Superficial acupuncture in the relief of chronic low back pain. *Ann R Coll Surg Engl* 1983;65(1):44-6.
73. Pariente J, White P, Frackowiak RS, Lewith G. Expectancy and belief modulate the neuronal substrates of pain treated by acupuncture. *Neuroimage* 2005;25(4):1161-7.
74. Birch S, Ida J. *Japanese Acupuncture*. Brookline: Paradigm Publications, 1998.
75. Miller FG, Kaptchuk TJ. Acupuncture trials and informed consent. *J Med Ethics* 2007;33(1):43-4.
76. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;328(7437):432.
77. Altman DG, Schulz KF, Moher D. Turning a blind eye: testing the success of blinding and the CONSORT statement. *BMJ* 2004;328(7448):1135; author reply 36.
78. Berman B, Gilpin AMK, Lao L. Acupuncture and Knee Osteoarthritis. *Ann Intern Med* 2005;142(10):872-3.
79. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol* 2007;36(3):664-5.
80. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;359(9307):696-700.

81. Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;328(7448):1136.
82. White P, Lewith G, Prescott P, Conway J. Acupuncture versus placebo for the treatment of chronic mechanical neck pain: a randomized, controlled trial. *Ann Intern Med* 2004;141(12):911-9.
83. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet* 1998;352(9125):364-5.
84. Ulett GA, Han S, Han JS. Electroacupuncture: mechanisms and clinical application. *Biol Psychiatry* 1998;44(2):129-38.
85. Brinkhaus B, Witt CM, Jena S, Linde K, Streng A, Wagenpfeil S, Irnich D, Walther HU, Melchart D, Willich SN. Acupuncture in patients with chronic low back pain: a randomized controlled trial. *Arch Intern Med* 2006;166(4):450-7.
86. Linde K. [Should acupuncture be reimbursed? Three acupuncture programmes of German Statutory Health Insurance Funds (SHI) and their scientific evaluation]. *Z Arztl Fortbild Qualitatssich* 2004;98(6):467-8.
87. Ernst E. Testing acupuncture for osteoarthritis: pragmatic trials or efficacy studies? *Rheumatology (Oxford)* 2006;45(2):125.
88. Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, Kirsch I, Schyner RN, Nam BH, Nguyen LT, et al. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *BMJ* 2008;336(7651):999-1003.
89. Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 2005;330(7501):1202-5.
90. Kaptchuk TJ. The placebo effect in alternative medicine: can the performance of a healing ritual have clinical significance? *Ann Intern Med* 2002;136(11):817-25.
91. Linde K, Witt CM, Streng A, Weidenhammer W, Wagenpfeil S, Brinkhaus B, Willich SN, Melchart D. The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain. *Pain* 2007;128(3):264-71.
92. Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, Stone DA, Nam BH, Kirsch I, Goldman RH. Sham device v inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006;332(7538):391-7.
93. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 2001;344(21):1594-602.
94. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998;316(7127):285.
95. Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;86(3):217-25.
96. Bjordal JM, Klovning A, Ljunggren AE, Slordal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: A meta-analysis of randomised placebo-controlled trials. *Eur J Pain* 2007;11(2):125-38.
97. Fransen M, McConnell S, Bell M. Exercise for osteoarthritis of the hip or knee. *Cochrane Database Syst Rev* 2003(3):CD004286.
98. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. *Osteoarthritis Cartilage* 2007;15(9):981-1000.
99. Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. *Ann Intern Med* 2002;136(6):471-6.
100. Suarez-Almazor M, Looney C, Street JR, Liu Y, Cox V, Pietz K, Marcus D. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of provider communication style. 2007;007-2. Osteoarthritis: Clinical Aspects I:Abstract No. 739.

Chapter 9. The effects of acupuncture on rates of clinical pregnancy among women undergoing in vitro fertilization: systematic review and meta-analysis

Manheimer E, van der Windt D, Cheng K, Stafford K, Liu J, Tierney J, Lao L, Berman BM, Langenberg P, Bouter LM. The effects of acupuncture on rates of clinical pregnancy among women undergoing in vitro fertilization: a systematic review and meta-analysis. Hum Reprod Update 2013 June 27 [Epub ahead of print].

ABSTRACT

BACKGROUND: Recent systematic reviews of adjuvant acupuncture for in vitro fertilization (IVF) have pooled heterogeneous trials, without examining variables that might explain the heterogeneity. The aims of our meta-analysis were to quantify the overall pooled effects of adjuvant acupuncture on IVF clinical pregnancy success rates, and evaluate whether study design-, treatment-, and population-related factors influence effect estimates.

METHODS: We included randomized controlled trials that compared needle acupuncture administered within one day of embryo transfer, versus sham acupuncture or no adjuvant treatment. Our primary outcome was clinical pregnancy rates. We obtained from all investigators additional methodological details and outcome data not included in their original publications. We analyzed sham-controlled and no adjuvant treatment-controlled trials separately, but since there were no large or significant differences between these two subsets, we pooled all trials for subgroup analyses. We pre-specified 11 subgroup variables (5 clinical and 6 methodological) to investigate sources of heterogeneity, using single covariate meta-regressions.

RESULTS: Sixteen trials (4,038 participants) were included in the meta-analyses. There was no statistically significant difference between acupuncture and control when combining all trials (risk ratio (RR) 1.12, 95% CI 0.96-1.31; I^2=68%; 16 trials; 4,021 participants), or when restricting to sham-controlled (RR 1.02, 0.83-1.26; I^2=66%; 7 trials; 2,044 participants) or no adjuvant treatment-controlled trials (RR 1.22, 0.97-1.52; I^2=67%; 9 trials; 1,977 participants). The type of control used did not significantly explain the statistical heterogeneity (interaction p=0.27). Baseline pregnancy rate, measured as the observed rate of clinical pregnancy in the control group of each trial, was a statistically significant effect modifier (interaction p<0.001), and this covariate explained most of the heterogeneity of the effects of adjuvant acupuncture across all trials (adjusted R^2=93%; I^2 residual=9%). Trials with lower control group rates of clinical pregnancy showed larger effects of adjuvant acupuncture (RR 1.53, 1.28-1.84; 7 trials; 1,732 participants) than trials with higher control group rates of clinical pregnancy (RR 0.90, 0.80-1.01; 9 trials; 2,289 participants). The asymmetric funnel plot showed a tendency for the intervention effects to be more beneficial in smaller trials.

CONCLUSIONS: We found no pooled benefit of adjuvant acupuncture for IVF. The subgroup finding of a benefit in trials with lower, but not higher, baseline pregnancy rates was the only statistically significant subgroup finding in our earlier review, has been confirmed in this update, and was not explained by any confounding variables evaluated. However, this baseline pregnancy rate subgroup finding among published trials requires further confirmation and exploration in additional studies because of the multiple subgroup tests conducted, the risk of unidentified confounders, the multiple different factors that determine baseline rates, and the possibility of publication bias.

The online version of the thesis Table of Contents includes links to the full-text of the following related material:

- Supplementary Appendices;
- Preliminary version of review, published in *BMJ* in 2008, as well as responses to comments on the *BMJ* review:
    - Manheimer E, Zhang G, Udoff L, Haramati A, Langenberg P, Berman BM, Bouter LM. Effects of acupuncture on rates of pregnancy and live birth among women undergoing in vitro fertilisation: systematic review and meta-analysis. *BMJ* 2008 Mar 8;336(7643):545-9.
    - Manheimer E, Bouter LM. Does acupuncture improve the results of in vitro fertilization? [letter: authors' response to commentary by Ernst E] *Focus on Alternative Therapies (FACT)* 2008 Sept;13(3):195.
    - Manheimer E, Berman BM, Bouter LM. Effects of acupuncture on rates of pregnancy and live birth among women undergoing in vitro fertilization: systematic review and meta-analysis [letter: authors' response to *BMJ* rapid responses] *BMJ* [online only] 2008 May 2.
- Additional preliminary work:
    - Manheimer E. Selecting a control for in vitro fertilization and acupuncture randomized controlled trials (RCTs): how sham controls may unnecessarily complicate the RCT evidence base. *Fertil Steril* 2011 Jun 30;95(8):2456-61.
    - Stener-Victorin E*, Manheimer E*. Commentary on the Cochrane review of acupuncture and assisted conception. *Explore (NY)* 2011 Mar-Apr;7(2):120-133. *Both authors contributed equally to this work.
    - Manheimer E, van der Windt D, Stafford K, Liu J, Tierney J, Lao L, Berman B, Bouter L. Effects of adjuvant acupuncture on pregnancy rates among women undergoing in vitro fertilization: systematic review and meta-analysis. [Protocol] PROSPERO 2011:CRD42011001309.

INTRODUCTION

Some 10-15% of couples have difficulty conceiving at some point in their reproductive lives, and seek specialist fertility care.[1] One of the most commonly used treatment options is in-vitro fertilization (IVF) and related expansions of it (e.g., intracytoplasmic sperm injection (ICSI)[2]). In 2003, it was estimated that 932,000 IVF and related cycles were performed worldwide, resulting in an estimated 232,000 babies being born.[3] The United States has among the highest IVF success rates worldwide, which is likely due to a combination of factors, including few restrictions on the number of embryos that can be transferred and the routine use of embryo selection.[3,4] European countries generally have lower IVF success rates because they are increasingly moving towards single embryo transfers,[5,6] which carry lower health risks [7]; the rates of multiple births and its associated health risks for mother and baby in Europe are approximately half of those in the United States.[3]

Despite many recent technological advances, average success rates with IVF remain low, with only 30% of treatment cycles resulting in a live birth in the United States.[8] Consequently, there is a need to investigate new laboratory techniques and drug therapies to improve the success rates of IVF, by means other than increasing the number of embryos transferred. However, progress in developing such safe and effective therapies has been limited,[9] driving patients to consider IVF-adjuvant complementary and alternative medical (CAM) therapies,[10] many of which are unproven and inadequately investigated. Acupuncture is the most commonly used adjuvant CAM fertility treatment among couples seeking fertility care in US fertility clinics.[10,11] "Fertility problems" is the second most common health condition, following pain-related conditions, for which people seek acupuncture treatment in the United Kingdom.[12] A review of the effects of adjuvant acupuncture on IVF is warranted because acupuncture has been found to be relatively safe among general patient populations[13] and also among women at various stages of pregnancy[14-16]; it is low cost, and if effective in increasing IVF success rates, can potentially reduce the need for an additional high cost IVF cycle; and qualitative research suggests that adjuvant acupuncture may help IVF patients deal with the psychological and emotional issues that accompany both subfertility and IVF treatment.[17-20]

We updated our previous systematic review and meta-analysis[21] with nine new trials including 2,672 new randomized participants, and carried out an *a priori* defined set of subgroup analyses to evaluate whether estimates of the effects of adjuvant acupuncture on IVF success rates is influenced by study design-, treatment-, and population-related factors.

METHODS

The systematic review protocol was registered in the PROSPERO database of prospectively registered systematic reviews (identification number: CRD42011001309) in May 2011.[22] The systematic review follows PRISMA reporting guidelines (for the PRISMA checklist, see Supplementary Appendix Text 1).[23]

Identification of studies

We searched Medline (OVID), Embase (Elsevier Sciences), CENTRAL (2012, Issue 4) (OVID), and the Chinese databases Sino-Med (previously called the Chinese Biomedical Database), Chinese National Knowledge Infrastructure, and VIP Database for Chinese Technical Periodicals, from inception to April 2012 (for the MEDLINE search strategy, see Supplementary Appendix Text 2). We also searched the proceedings of the following three major annual conferences on assisted reproductive technology for 2001-12: American Society for Reproductive Medicine; European Society of Human Reproduction and Embryology; and Pacific Coast Reproductive Society. We also searched for previous systematic reviews on this topic, and reviewed their reference lists.[21,24-32] In addition, we sent the provisional list of included and excluded trials to experts in the field, and

asked if they were aware of any potentially eligible trials, published or unpublished, that were not on our list. To identify trials that may be relevant for future updates of this review, we also searched the following databases of ongoing trials: the US National Institutes of Health's clinicaltrials.gov, the World Health Organization's International Clinical Trials Registry Platform, and controlled-trials.com.

Selection criteria, data extraction, and quality assessment

We sought randomized controlled trials (RCTs) that compared acupuncture with sham acupuncture or no adjuvant treatment. Because we evaluated acupuncture as a complement to embryo transfer, we considered only RCTs in which acupuncture was administered within one day of the procedure, with the objective of improving IVF success rates. We excluded RCTs which evaluated electrical acupuncture as an alternative to conventional anaesthesia for the surgical procedure of removing the oocytes from the woman's ovaries. We considered these two sets of RCTs to be fundamentally different, in terms of the aim of the treatment with acupuncture, the timing of acupuncture administration (i.e., during oocyte retrieval versus embryo transfer) and the acupuncture protocol used. That is, when acupuncture is used during oocyte retrieval, the primary purpose is to relieve pain, and correspondingly, electrical stimulation of the needles is always used[33] and the points are selected to effect pain reduction. In contrast, when acupuncture is used to accompany embryo transfer, the primary purpose is to assist conception, electrical stimulation of the needles may or may not be used, and the points are selected to improve blood flow to the uterus to make it more receptive to the embryo.[34] In addition, two systematic reviews[35,36] have already evaluated electro-acupuncture as an alternative to anesthesia during ooctye retrieval, and pooled analyses found no statistically significant differences in later pregnancy rates between the electro-acupuncture and conventional anesthesia groups. Trials that included ICSI as part of the IVF procedure were eligible.

We included only trials in which acupuncture involved the insertion of needles into traditional meridian points. The needles could be inserted into tender points in addition to the traditional meridian points, and the needles could also be electrically stimulated.

We included RCTs in any language, published as either full articles or abstracts. We also included any unpublished trials that we identified, but only if we were able to obtain sufficient and reliable information on their methods and outcomes.

For trials to be eligible, we had to be able to extract data on at least one of the following outcomes, as recommended[37-39]: clinical pregnancy (that is, presence of at least one gestational sac or fetal heartbeat, confirmed by transvaginal ultrasound[40]), ongoing pregnancy (that is, pregnancy beyond 12 weeks of gestation, as confirmed by fetal heart activity on ultrasound), or live birth.

Two authors independently selected articles and extracted data, with disagreements resolved by discussion. We extracted data pertaining to quality of the methods, participants, interventions, and outcomes. We contacted corresponding authors with specific questions related to the design and outcomes of their trials and asked them to review the information we extracted from their trials and clarify any ambiguities.

We evaluated the methodological quality of the trials using the Cochrane risk of bias tool,[41] supplemented with an additional item that assessed whether there was a co-intervention that was applied unequally across treatment groups. We added 'unequal co-intervention' to the Cochrane tool because this item can indicate an important source of bias and has been included in other quality assessment tools,[42] including an assessment tool used in subfertility meta-analyses.[43]

Data synthesis and analysis

The pooled risk ratio of achieving a clinical pregnancy for women in the acupuncture group compared with women in the control group was the primary outcome measure. We pre-specified clinical pregnancy instead of live birth as our primary outcome because it would allow for more RCTs and more data to contribute to our primary analyses, as only 22% of subfertility RCTs that report the clinical pregnancy outcome data also report the live birth outcome data.[44] Therefore, including clinical pregnancy as the primary outcome will avoid the risk of including in the primary analysis only a subset of the RCTs (i.e., only those RCTs that report live births), which may not be representative of all RCTs included in the review.[21] In addition, clinical pregnancy is a practical and clinically relevant surrogate for live birth because approximately 84% of clinical pregnancies achieved through IVF would be expected to result in a live birth, based on data from IVF RCTs.[44] Also, the effect estimates on the clinical pregnancy outcome have been shown to be nearly identical to the effect estimates on the live birth outcome for IVF RCTs that report both outcomes (i.e., ratio of odds ratios: 0.99 (95% CI, 0.87 to 1.13)).[44] Finally, the clinical pregnancy outcome may allow for easier control of potentially confounding factors, which might be introduced after the first trimester, once the women are no longer under the care of the IVF clinic. That is, collecting reliable information about live births after patients have left the IVF clinic might be difficult because patients, who may travel long distances to reach the clinic, might be less accessible to the clinic staff after the clinical pregnancy is achieved.[37,38,44] Ongoing pregnancy, live birth, and miscarriage rates were analysed as secondary outcomes.

For pooled data, summary test statistics were calculated using the DerSimonian and Laird model,[45,46] which is the random effects model used in the RevMan software, version 5.1.[47] This model estimates the average treatment effect by incorporating heterogeneity among clinically diverse trials with different, but related, treatment effects. When heterogeneity exists, the model assigns smaller studies more weight than they would receive in a fixed-effects model.[48] We used the random effects model because of the expected heterogeneity of the studies' acupuncture protocols and settings. We evaluated heterogeneity using both the I^2 statistic, which indicates the proportion of variability across trials not explained by sampling variation alone, and the p-value of the chi^2 test of heterogeneity.[48] Although interpreting the importance of inconsistency depends on other factors in addition to the I^2 values (e.g., p-value from chi^2 test, magnitude and direction of effects), the Cochrane Handbook suggests the following rough guide to interpreting I^2 values: "0 to 40% might not be important; 30% to 60% may represent moderate heterogeneity; 50% to 90% may represent substantial heterogeneity; and 75% to 100% may represent considerable heterogeneity".[48]

We included in the meta-analyses all randomized women who began the IVF process including those who did not complete the treatment (that is, had no embryo transfer), even if these participants were excluded in the author's analysis of the trial. Although inclusion of women without an embryo transfer will tend to underestimate the effect of acupuncture,[38] it is the more conservative and appropriate analytical approach[38,39,49,50] because it preserves the groups created by the randomization and reduces the chance of a type I error.[38] The only randomized participants we did not include were those for whom the clinical pregnancy or live birth outcome data were missing.

We analysed the sham-controlled and no adjuvant treatment-controlled trials separately, but if there was no large or significant difference in pooled effect estimates between these two subgroups of trials, we pooled all available trials.

If at least ten trials were available for a meta-analysis, we assessed for the likelihood of publication bias by constructing funnel plots.[51]

Subgroup analyses

We conducted subgroup analyses on five clinical characteristics that might influence the effect of adjuvant acupuncture on clinical pregnancy success rates: 1) two acupuncture sessions or more than two; 2) selection of meridian acupuncture points the same as the points selected in the first published trial[52] that evaluated acupuncture as an adjuvant to embryo transfer, and which showed a large effect, or a modified version of this trial's acupuncture point selection protocol; 3) control group clinical pregnancy rate (as an estimate of the baseline clinical pregnancy rate) dichotomized as higher (32% or greater, which is the European average of pregnancy rate per embryo transfer[53]) or lower; control group clinical pregnancy rate was also analyzed as a continuous variable to test whether the relation was linear and consistent with the findings of the categorical analysis; 4) explanatory trials conducted to test the effects of adjuvant acupuncture under controlled conditions in which the acupuncture was administered onsite at the IVF clinic or pragmatic trials conducted to test the effects of adjuvant acupuncture delivered offsite, which might better approximate every day, "real life" conditions since most IVF clinics do not have onsite acupuncturists[37]; and 5) trials that involved a treating acupuncturist who was judged as adequately experienced or not adequately experienced, with such judgments made by acupuncturist assessors who were blinded to the identities and results of the trials.

In addition, to assess whether the effects varied with the risk of bias domains of the trials, we also conducted subgroup analyses[48] on the following six 'risk of bias' domains: random allocation sequence generation; concealment of allocation of randomization sequence; blinding of patients (i.e., use of sham control); blinding of embryo transfer physicians; incomplete outcome data; and unequal co-intervention.

For each subgroup analysis, we performed a single covariate weighted random effects meta-regression[54-56] in Stata version 11 (StataCorp)[57] to investigate whether differences in effects of adjuvant acupuncture between the covariate's two subgroups were statistically significant. In interpreting the importance of subgroup effects, we also considered the difference in the magnitude or direction of effect between the two subgroups as well as whether or not the confidence intervals of the subgroups overlapped.[48] For each single covariate meta-regression subgroup analysis, we calculated the p-value of the test for interaction; the percentage of the between-study variance explained by the covariate (adjusted R^2); and the percentage of the residual variation that is attributable to between study heterogeneity (I^2_{res}). A random effects meta-regression model was used rather than a fixed effects model or univariate subgroup testing because the random effects meta-regression model allows for potential residual heterogeneity and it is therefore a more conservative analytical approach. We did not attempt to develop a multivariate meta-regression model, primarily because of the relatively small number of studies relative to the number of subgroup variables pre-specified, but also because our objective was to identify modifiers of the effects of adjuvant acupuncture on IVF, rather than to build an optimal 'prediction' model. A major limitation of subgroup analyses is the potential bias by confounding.[54] Therefore, we also examined whether the apparent effect modification due to a specific subgroup variable might be explained by an association between that subgroup variable of interest and a potentially confounding subgroup variable, by including both variables in a meta-regression model.

Sensitivity analyses

For our primary analyses, we excluded participants with unrecorded clinical pregnancy data; however, for sensitivity analyses, we imputed the rate of clinical pregnancies in participants with unrecorded outcomes using available case rates: 1) from the corresponding trial group that these participants were assigned to (i.e., either acupuncture or control); and 2) from the corresponding trial's acupuncture plus control group combined. We also conducted a third sensitivity analysis which assumed that no randomized participants with an unrecorded outcome achieved a clinical pregnancy.

RESULTS

Supplementary Appendix Figure 1 shows the study selection process. The list of studies that were excluded with the reasons for exclusion, and the list of ongoing trials and their characteristics are available in Supplementary Appendix Tables 1 and 2, respectively.

Sixteen RCTs with a total of 4,038 participants met the inclusion criteria (Table 1; see Supplementary Appendix Table 3 for full details). All trials were published in English since 2002, and conducted in seven different countries. Ten were published as full reports[52,58-66] and six as abstracts.[67-72] We obtained unpublished methodological information for all trials, and unpublished live birth outcome data for 7 trials.[52,59,66-68,71,72]

Trial characteristics

The only differences in trial eligibility criteria were that two German trials[52,72] included only women with good quality embryos whereas the other trials included women with embryos of varying quality; one trial[67] restricted eligibility to women with an unfavourable reproductive prognosis; one trial[65] restricted eligibility to frozen-thawed embryo transfer cycles while the others used fresh embryos; and one trial[61] used ICSI for all participants, whereas all other trials reported use of ICSI for only some participants.

The timing of the acupuncture sessions relative to embryo transfer differed somewhat among trials (Table 1). In all trials, however, women received acupuncture on the day of embryo transfer, before and/or after the embryo transfer procedure. In all but one trial,[68] the acupuncture was administered directly on-site of the IVF clinic. The number of acupuncture treatments ranged from one to three. In all trials, the acupuncture sessions lasted 25-30 minutes. In all trials, the acupuncture protocol and selection of acupuncture points was designed for the sole purpose of improving rates of pregnancy.

Twelve trials reported that the "de qi" needling sensation (i.e., a pain, achiness, stinging, or dullness at the needle insertion site, which is an indicator that the acupuncture needle has been correctly placed) was sought,[52,59,61-69,72] whereas one sham-controlled trial reported that there was no attempt to manipulate the needles to achieve the "de qi" needling sensation, in order to avoid unblinding trial participants.[58] The three other trials did not report on de qi.[60,70,71] No trial used electroacupuncture. For all trials, the mean numbers of embryos transferred was similar between the randomized groups.

Table 1. Characteristics of included trials

Study, Country [reference]	N*	Acupuncture	Control type†
Andersen, Denmark[58]	635	2 sessions: 1) 30 m before ET; 2) immediately after ET. [To maintain blinding, needles not manipulated to obtain de qi needling sensation]; (n=314)	Non-penetrating blunt needles, placed on the real acupoints [and not manipulated to achieve de qi sensation] (n=321)
Arnoldi, Italy[67]	204	3 sessions: 1) d 5 of ovarian stimulation; 2) 30 m before ET; 3) immediately after ET. [Needles stimulated manually by rotating, lifting and thrusting the handle of the needle in order to maintain de qi sensation, both during initial insertion and after 10 m]; (n=102)	No adjuvant treatment (n=102)
Craig, US[68]	[113]	2 sessions, both at an off-site location: 1) within 1 to 2 h before ET; 2) within 1 to 2 h after ET. [Needles simulated manually at insertion, and then manually rotated after 10 m to maintain de qi] (n=57).	No adjuvant treatment (n=56)
Dieterle, Germany[59]	225	2 sessions: 1) immediately after ET; 2) 3 days after ET. Needles rotated at the start and after 15 m, to evoke de qi; (n=116)	Needle acupuncture, with needles inserted in real acupoints not expected to influence fertility (n=109) ‡
Domar, US[60]	150	2 sessions: 1) 25 m before ET; 2) immediately after ET. (n=78) §	No adjuvant treatment (n=68)
Feliciani, Italy[69]	46	3 sessions: 1) 5-7 d before egg retrieval; 2) 2-3 d before egg retrieval; 3) within 1 h after ET. [Needles manipulated to obtain de qi]; (n=23)	No adjuvant treatment (n=23)
Fratterelli, US[70]	397	2 sessions: 1) 25 m before ET; 2) after ET. (n=200) §	No adjuvant treatment (n=197) ‖
Madaschi, Brazil[61]	[455] ¶	2 sessions: 1) 25 m before ET; 2) immediately after ET. Needles manipulated until de qi sensation obtained; (n=230)	No adjuvant treatment (n=225)
Moy, US[62]	161	2 sessions: 1) 25 m before ET; 2) immediately after ET. [Needles rotated to obtain de qi sensation]; (n=87)	Needle acupuncture, with body acupuncture insertion sites close to, but not on, the real acupoints, with same manipulation of needles, as performed in true acupuncture group (n=74)
Omodei, Italy[71]	168	2 sessions: 1) 25 m before ET; 2) immediately after ET. (n=44) §	No adjuvant treatment (n=124)
Paulus 2002, Germany[52]	160	2 sessions: 1) 25 m before ET; 2) immediately after ET. Needles rotated at the start and after 10 m, to evoke de qi; (n=80)	No adjuvant treatment (n=80)
Paulus 2003, Germany[72]	200	2 sessions: 1) 25 m before ET; 2) immediately after ET. Needles rotated after 10 m to evoke de qi; (n=100)	Non-penetrating blunt needles, placed on the real acupoints (n=100) ‡
Smith, Australia[63]	228	3 sessions: 1) d 9 of stimulating injections; 2) 25 m before ET; 3) immediately after ET. [Needles manually stimulated to obtain de qi sensation]; (n=110)	Non-penetrating blunt needles, placed close to, but not on, the real acupoints, with the sham needles manipulated by lifting and thrusting the handle of the needles and running a fingernail along the handle, but de qi not sought (n=118)
So 2009, Hong Kong/China[64]	370	2 sessions: 1) 25 m before ET; 2) immediately after ET. Needles stimulated manually by rotating, lifting and thrusting the handle of the needle in order to maintain de qi sensation, both during initial insertion and after 10 m; (n=185)	Non-penetrating blunt needles, placed on the real acupoints, and manipulated in the same way as the true acupuncture needles to give patients a pricking, penetrating sensation (n=185)

Study, Country [reference]	N*	Acupuncture	Control type†
So 2010, Hong Kong/China[65]	226	1 session immediately after ET. Needles stimulated manually by rotating, lifting and thrusting the handle of the needle in order to maintain de qi sensation, both during initial insertion and after 10 m; (n=113)	Non-penetrating blunt needles, placed on the real acupoints, and manipulated in the same way as the true acupuncture needles to give patients a pricking, penetrating sensation (n=113)
Westergaard, Denmark[66]	300	Acupuncture group 1: 1) 25 m before ET; 2) immediately after ET (n=100); Acupuncture group 2: 1) 25 m before ET; 2) immediately after ET; 3) 2 d after ET. Needles rotated at the start and after 10 m, to evoke de qi; (n=100) **	No adjuvant treatment (n=100)

Additional data obtained from RCT authors are enclosed in brackets to allow such data to be differentiated from the data included only in the publications. Acupoints, Acupuncture points; d, day; ET, embryo transfer; h, hour; m, minute.

* Number randomized.

† For all sham controlled trials, the sham acupuncture procedure was given on the same schedule as that used for the true acupuncture group.

‡ These two trials did not report whether the sham needles were manipulated in the same way as the true acupuncture needles.

§ These three trials did not report whether or not the needles used in the true acupuncture group were manipulated to achieve the de qi sensation.

|| We excluded from our meta-analysis the following 3 additional control groups included in this trial: laser acupuncture, sham laser acupuncture, and relaxation.

¶ The 39 participants (22 acupuncture; 17 control) who did not proceed to embryo transfer were excluded from the trial authors' analysis. The approach for our meta-analysis was to re-include these participants (i.e., acupuncture group = 230 (i.e., 208 +22) and the no acupuncture group = 225 (i.e., 208 + 17)), because it could be assumed that these participants without an embryo transfer did not achieve the clinical pregnancy outcome.

** For this trial, acupuncture groups 1 and 2 were combined together for the meta-analysis.

Methodological quality of included trials

For one trial,[61] the method of generating the random allocation sequence was not clear; however, the randomization appeared to be successful in creating similar groups, as there were no baseline differences in prognostic factors between treatment groups. For the other 15 trials, adequate methods of sequence generation were used. For three trials,[61,67,69] there was inadequate allocation concealment (Table 2; see Supplementary Appendix Table 4 for full details), but in these trials there was also baseline similarity between the two groups. In three trials,[62,66,68] there were small numbers of randomized women with missing clinical pregnancy outcomes (Figure 1, footnote). In four trials,[61,63,66,68] some randomized women began the IVF process but did not complete the treatment (that is, no embryo transfer); however, as noted above, these women were still included in the meta-analyses.

Seven of the trials used a sham acupuncture control,[58,59,62-65,72] of which four[58,64,65,72] used the Streitberger non-penetrating sham needles[73,74] placed on the true acupuncture points (Table 1). The remaining nine trials used a no adjuvant treatment control. For ten trials, the embryo transfer physicians were blinded to the treatment assignments [52,58-62,64,65,68,72]; for the other six trials the physicians were not blinded[63,66,67,69-71] (Table 2).

Table 2. Risk of bias summary of included trials

Study, Country [reference]	Allocation sequence generation	Allocation concealment	Blinding of patients	Blinding of physicians	Incomplete outcome data addressed: Clinical pregnancy	Selective reporting	Unequal co-interven-tion
Andersen, Denmark[58]	Low risk	[Low risk]	Low risk	Low risk	Low risk	Low risk	Low risk
Arnoldi, Italy[67]	Low risk	[High risk*]	High risk	[High risk]	[Low risk†]	Low risk	Low risk
Craig, US[68]	[Low risk]	[Low risk]	High risk	Low risk	[Low risk]	Low risk	High risk
Dieterle, Germany[59]	[Low risk]	[Low risk]	Low risk	Low risk	Low risk	Low risk	Low risk
Domar, US[60]	Low risk	[Low risk]	High risk	[Low risk]	[Low risk‡]	Low risk	Low risk
Feliciani, Italy[69]	[Low risk]	[High risk]	High risk	[High risk]	[Low risk]	Low risk	Low risk
Fratterelli, US[70]	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk
Madaschi, Brazil[61]	[Unclear§]	[High risk‖]	High risk	[Low risk]	[Low risk¶]	Low risk	Low risk
Moy, US[62]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Omodei, Italy[71]	[Low risk]	[Low risk]	High risk	[High risk]	[Low risk]	Low risk	Low risk
Paulus 2002, Germany[52]	Low risk	[Low risk]	High risk	Low risk	[Low risk]	Low risk	Low risk
Paulus 2003, Germany[72]	[Low risk]	[Low risk]	Low risk	[Low risk]	[Low risk]	Low risk	Low risk
Smith, Australia[63]	[Low risk]	[Low risk]	Low risk	[High risk]	[Low risk]	Low risk	Low risk
So 2009, Hong Kong/China[64]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
So 2010, Hong Kong/China[65]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Westergaard, Denmark[66]	[Low risk]	[Low risk**]	High risk	High risk	[Low risk]	Low risk	Low risk

Additional data obtained from RCT authors are enclosed in brackets to allow such data to be differentiated from data included only in the publications.

* Although sealed envelopes were used, the envelopes were not sequentially numbered and the trial's investigators could not recall whether or not the envelopes were opaque.

† Although there was a large imbalance in the number of women who did not proceed to an embryo transfer (i.e., 6/102 in acupuncture group and 20/102 in control group), which the trial authors postulated to be due to the acupuncture increasing the likelihood of a viable embryo being available for transfer, the outcomes for the randomized participants without embryo transfer were known (i.e., not pregnant), so we did not consider this as a bias due to incomplete outcome data.

‡ The treatment assignment and the outcomes for 4/150 randomized participants were not recorded by the trial authors. Therefore, the 146 participants analyzed were instead used in the authors' analysis,[60] and for this meta-analysis. However, we scored this criterion as "low risk" for incomplete outcome data because the reasons for missing outcome data were unlikely to be related to the outcomes and the proportion of missing outcomes was not likely to have a clinically relevant impact on the effect estimates.[41]

§ A computer-generated randomization list was used to assign patients to treatment groups. However, this trial was judged as unclear for "allocation sequence generation" because it was not clear how the trial authors assigned to treatment groups 39 new participants who replaced the 39 participants excluded because of no embryo transfer.

‖ This trial used an open randomization list, so the investigators enrolling participants could possibly foresee assignments and thus introduce selection bias.

¶ In this trial, the randomization occurred prior to the start of the ovarian stimulation, although the participants were not informed of their treatment assignment until the start of the ovarian stimulation. Approximately 15% of participants were randomized, but then decided not to get IVF, primarily because of its costs, and withdrew prior to the start of the ovarian stimulation. Because these participants withdrew from the study before they were told whether they had been randomized to acupuncture or control, their decision to withdraw from the trial could not have been affected by knowledge of the randomized intervention (i.e., acupuncture or control), and these withdrawals would not be expected to cause an important bias due to missing outcome data.[41]

** For this trial, the randomization treatment assignments were placed in sealed, opaque envelopes, which were shuffled and deposited in a cardboard box, from which each participant selected only one. This procedure has handled by an independent nurse not responsible for obtaining information about patients and enrolling them. Although the envelopes were not sequentially numbered, we considered the safeguards used in the randomization process to have provided adequate assurance of allocation concealment.

Efficacy analysis

There were no statistically significant pooled benefits of adjuvant acupuncture relative to either control, for the clinical pregnancy, ongoing pregnancy, or live birth outcomes; however, the overall statistical heterogeneity was substantial for each outcome (I^2=68%, 69%, 68%; chi^2 p-value=0.0001, 0.0004, and 0.0003, respectively) (Figure 1). The type of control used did not significantly explain this statistical heterogeneity (interaction p=0.27, for clinical pregnancy outcome) (Figure 1).

Of the 11 variables planned for subgroup analyses for the clinical pregnancy outcome, only 10 could be tested because no trial had a high risk of bias from incomplete outcome data. In addition, only the Craig et al trial[68] used an 'unequal co-intervention', which was driving to and from the off-site acupuncturist's office, before and after the embryo transfer procedure, in only the acupuncture group. Therefore, the 'unequal co-intervention' and 'site of acupuncture administration' variables classified trials in the same subgroups, and therefore had the same subgroup effect results.

We combined all 16 trials for the meta-regression subgroup analyses. Of the 10 subgroup variables tested, only baseline clinical pregnancy rate was a statistically significant effect modifier of adjuvant acupuncture (Figure 2). Baseline pregnancy rate was also the only variable tested that explained a large proportion of the between study variance. That is, in the meta-regression model fit with the single covariate of baseline clinical pregnancy rate, the adjusted R^2=90% when baseline clinical pregnancy rate was operationalized as a continuous variable (interaction p<0.001) and the adjusted R^2=93% when the baseline pregnancy rate was operationalized as a dichotomous variable (interaction p<0.001). To evaluate whether any of the other pre-specified subgroup variables was a confounder of the effects of baseline pregnancy on IVF clinical pregnancy success rates, we also fit nine other separate meta-regression models which included both baseline pregnancy rate as a continuous variable and one of each of the nine other pre-specified subgroup variables. In none of these nine other models did the correlation coefficient of the baseline clinical pregnancy rate variable substantially change (i.e., the maximum percentage change of the coefficient was 7%). To investigate whether the effect modification of baseline pregnancy rate was maintained across trial subsets, we also fit six separate single covariate (i.e., baseline pregnancy rate) meta-regression models, restricting to subsets of trials, according to type of control (sham or no adjuvant treatment), blinding of embryo transfer physician (yes or no), and allocation concealment (yes or no). In these meta-regression models, the magnitude of the baseline pregnancy rate subgroup effect was maintained when restricting to sham-controlled trials (76% difference in risk ratio estimates between sham-controlled trials with higher (RR=0.89), compared to sham-controlled trials with lower (RR=1.65), baseline pregnancy rates; interaction p=0.022; adj R^2=100%); no adjuvant treatment controlled trials (63% difference in RR; p=0.012; adj R^2=79%); trials with embryo transfer blinding (60% difference in RR; p=0.005; adj R^2=92%); trials without embryo transfer blinding (59% difference in RR; p=0.039; adj R^2=100%); trials with adequate allocation concealment (76% difference in RR; p<0.001; adj R^2=98%); and trials without adequate allocation concealment (40% difference in RR; p=0.661; adj R^2 could not be calculated because only three trials were included in this meta-regression model).

Figure 3 shows a graph of the meta-regression line, fit with the single continuous covariate of baseline pregnancy rate. This meta-regression model showed that a 20% and 40% increase in baseline pregnancy rate was associated, respectively, with a 33% and 55% decrease in the risk ratio of clinical pregnancy with adjuvant acupuncture. A 20% and 40% decrease in baseline clinical pregnancy rate was associated, respectively, with a 49% and 122% increase in the risk ratio of clinical pregnancy.

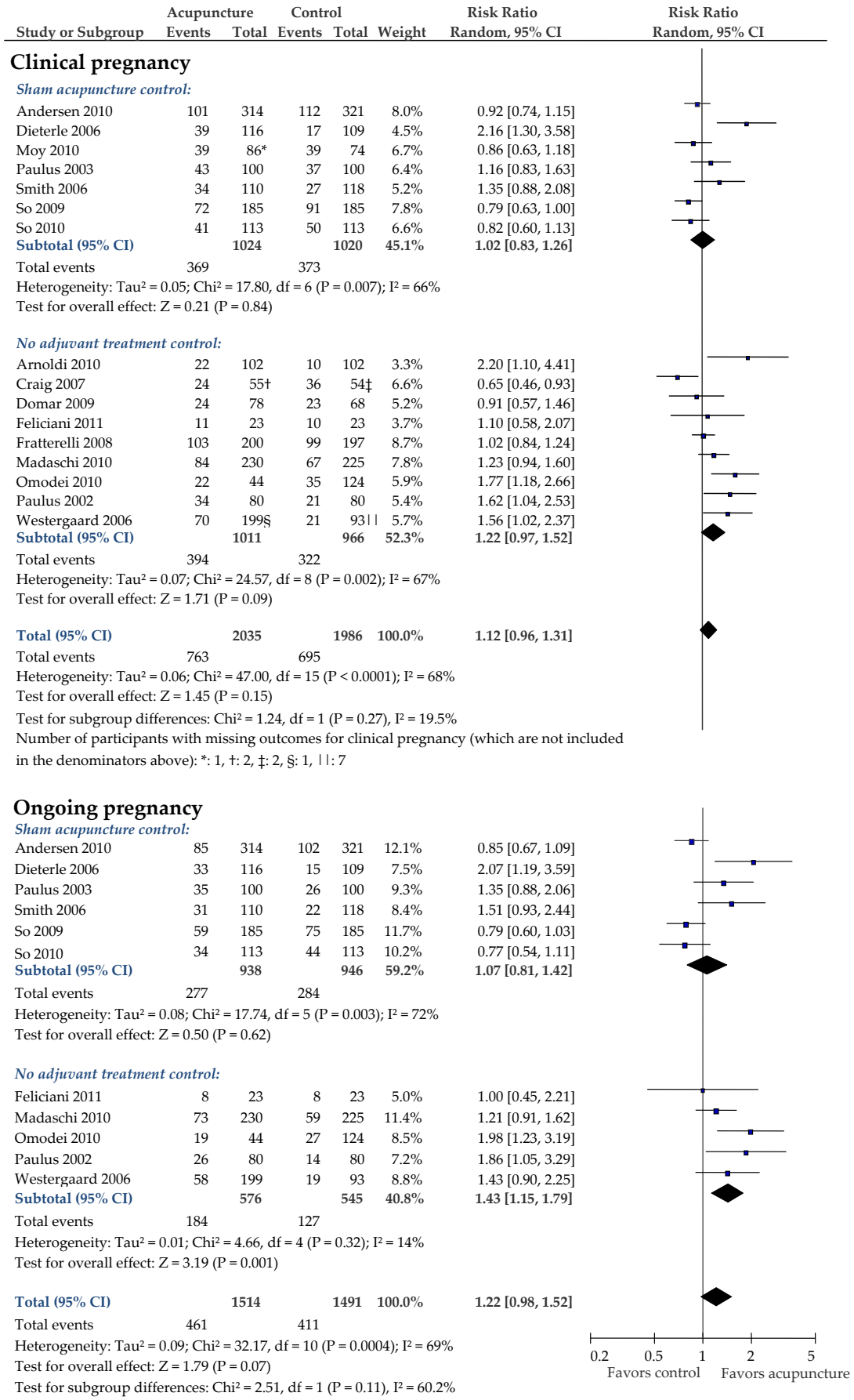
The subgroup variables 'blinding of embryo transfer physician' and 'site of acupuncture administration' were borderline statistically significant across all trials (interaction p=0.084 and 0.096, respectively); however, neither variable explained a large proportion of the between trial

variability (adj R^2=22% and 19%, respectively). Because the trials that blinded the embryo transfer physician also tended to blind the patients (i.e., use a sham control) (see Table 2), we were concerned about confounding between these two variables. Therefore we also fit a meta-regression model which included the patient blinding variable in addition to the embryo transfer physician blinding variable. In this model which included both variables, the correlation coefficient of the embryo transfer physician blinding variable was not substantially changed (i.e., from -0.158 to -0.150), and the interaction p-value of the embryo transfer blinding covariate only increased slightly (i.e., from p=0.084 to p=0.15). Finally, because the only trial that evaluated acupuncture administered off-site of the IVF clinic[68] used a no adjuvant treatment control, we also fit a meta-regression model using the single covariate of site of acupuncture administration, but restricting to the no adjuvant treatment controlled trial subset. In this meta-regression model, the site of acupuncture administration covariate was statistically significant (interaction p=0.04), and this covariate explained some of the heterogeneity within this subset of trials (adj R^2=61%). The embryo transfer blinding variable was not a statistically significant effect modifier when restricting to the sham-controlled (interaction p=0.476) or no adjuvant treatment-controlled (p=0.246) trial subsets. Although we initially planned subgroup analyses stratified by control type, we combined all trials for our final subgroup analyses shown in Figure 2 because of the largely similar subgroup effects across the two control groups.

The funnel plot (Figure 4) showed small study effects, with the intervention effects estimated in smaller studies showing more benefit than the effects estimated in larger studies. The Egger statistical test for funnel plot asymmetry found that the association between estimated intervention effects and the standard error of the intervention effect (as a measure of the trial's size) was greater than would be expected to occur by chance (p=0.032).

Sensitivity analyses

The use of the different methods of imputing the unrecorded outcome data values did not affect the results of the individual trials, any of the pooled results, or the results of any of the subgroup analyses. In trials that had participants with unrecorded clinical pregnancy outcomes,[62,66,68] imputing these values using available case rates from the corresponding trial group that these participants were assigned to resulted in very similar pooled estimates for the primary outcome measure (RR= 1.12; 95% CI, 0.96 to 1.30; I^2=68%; 16 trials). Imputing the values using the corresponding trial's acupuncture plus control group combined resulted in identical pooled estimates. Assuming that no participants with an unrecorded outcome achieved a clinical pregnancy resulted in very similar pooled estimates (RR= 1.13; 95% CI, 0.96 to 1.31; I^2=69%; 16 trials).


Acupuncture
Control
Risk Ratio
Risk Ratio
Study or Subgroup
Events
Total
Events
Total
Weight
Random, 95% CI
Random, 95% CI
Clinical pregnancy
Sham acupuncture control:
Andersen 2010 101 314 112 321 8.0% 0.92 [0.74, 1.15]
Dieterle 2006 39 116 17 109 4.5% 2.16 [1.30, 3.58]
Moy 2010 39 86* 39 74 6.7% 0.86 [0.63, 1.18]
Paulus 2003 43 100 37 100 6.4% 1.16 [0.83, 1.63]
Smith 2006 34 110 27 118 5.2% 1.35 [0.88, 2.08]
So 2009 72 185 91 185 7.8% 0.79 [0.63, 1.00]
So 2010 41 113 50 113 6.6% 0.82 [0.60, 1.13]
Subtotal (95% CI) 1024 1020 45.1% 1.02 [0.83, 1.26]
Total events 369 373
Heterogeneity: Tau² = 0.05; Chi² = 17.80, df = 6 (P = 0.007); I² = 66%
Test for overall effect: Z = 0.21 (P = 0.84)
No adjuvant treatment control:
Arnoldi 2010 22 102 10 102 3.3% 2.20 [1.10, 4.41]
Craig 2007 24 55† 36 54‡ 6.6% 0.65 [0.46, 0.93]
Domar 2009 24 78 23 68 5.2% 0.91 [0.57, 1.46]
Feliciani 2011 11 23 10 23 3.7% 1.10 [0.58, 2.07]
Fratterelli 2008 103 200 99 197 8.7% 1.02 [0.84, 1.24]
Madaschi 2010 84 230 67 225 7.8% 1.23 [0.94, 1.60]
Omodei 2010 22 44 35 124 5.9% 1.77 [1.18, 2.66]
Paulus 2002 34 80 21 80 5.4% 1.62 [1.04, 2.53]
Westergaard 2006 70 199§ 21 93|| 5.7% 1.56 [1.02, 2.37]
Subtotal (95% CI) 1011 966 52.3% 1.22 [0.97, 1.52]
Total events 394 322
Heterogeneity: Tau² = 0.07; Chi² = 24.57, df = 8 (P = 0.002); I² = 67%
Test for overall effect: Z = 1.71 (P = 0.09)
Total (95% CI) 2035 1986 100.0% 1.12 [0.96, 1.31]
Total events 763 695
Heterogeneity: Tau² = 0.06; Chi² = 47.00, df = 15 (P < 0.0001); I² = 68%
Test for overall effect: Z = 1.45 (P = 0.15)
Test for subgroup differences: Chi² = 1.24, df = 1 (P = 0.27), I² = 19.5%
Number of participants with missing outcomes for clinical pregnancy (which are not included in the denominators above): *: 1, †: 2, ‡: 2, §: 1, ||: 7
Ongoing pregnancy
Sham acupuncture control:
Andersen 2010 85 314 102 321 12.1% 0.85 [0.67, 1.09]
Dieterle 2006 33 116 15 109 7.5% 2.07 [1.19, 3.59]
Paulus 2003 35 100 26 100 9.3% 1.35 [0.88, 2.06]
Smith 2006 31 110 22 118 8.4% 1.51 [0.93, 2.44]
So 2009 59 185 75 185 11.7% 0.79 [0.60, 1.03]
So 2010 34 113 44 113 10.2% 0.77 [0.54, 1.11]
Subtotal (95% CI) 938 946 59.2% 1.07 [0.81, 1.42]
Total events 277 284
Heterogeneity: Tau² = 0.08; Chi² = 17.74, df = 5 (P = 0.003); I² = 72%
Test for overall effect: Z = 0.50 (P = 0.62)
No adjuvant treatment control:
Feliciani 2011 8 23 8 23 5.0% 1.00 [0.45, 2.21]
Madaschi 2010 73 230 59 225 11.4% 1.21 [0.91, 1.62]
Omodei 2010 19 44 27 124 8.5% 1.98 [1.23, 3.19]
Paulus 2002 26 80 14 80 7.2% 1.86 [1.05, 3.29]
Westergaard 2006 58 199 19 93 8.8% 1.43 [0.90, 2.25]
Subtotal (95% CI) 576 545 40.8% 1.43 [1.15, 1.79]
Total events 184 127
Heterogeneity: Tau² = 0.01; Chi² = 4.66, df = 4 (P = 0.32); I² = 14%
Test for overall effect: Z = 3.19 (P = 0.001)
Total (95% CI) 1514 1491 100.0% 1.22 [0.98, 1.52]
Total events 461 411
Heterogeneity: Tau² = 0.09; Chi² = 32.17, df = 10 (P = 0.0004); I² = 69%
Test for overall effect: Z = 1.79 (P = 0.07)
Test for subgroup differences: Chi² = 2.51, df = 1 (P = 0.11), I² = 60.2%
0.2
0.5
1
2
5
Favors control
Favors acupuncture

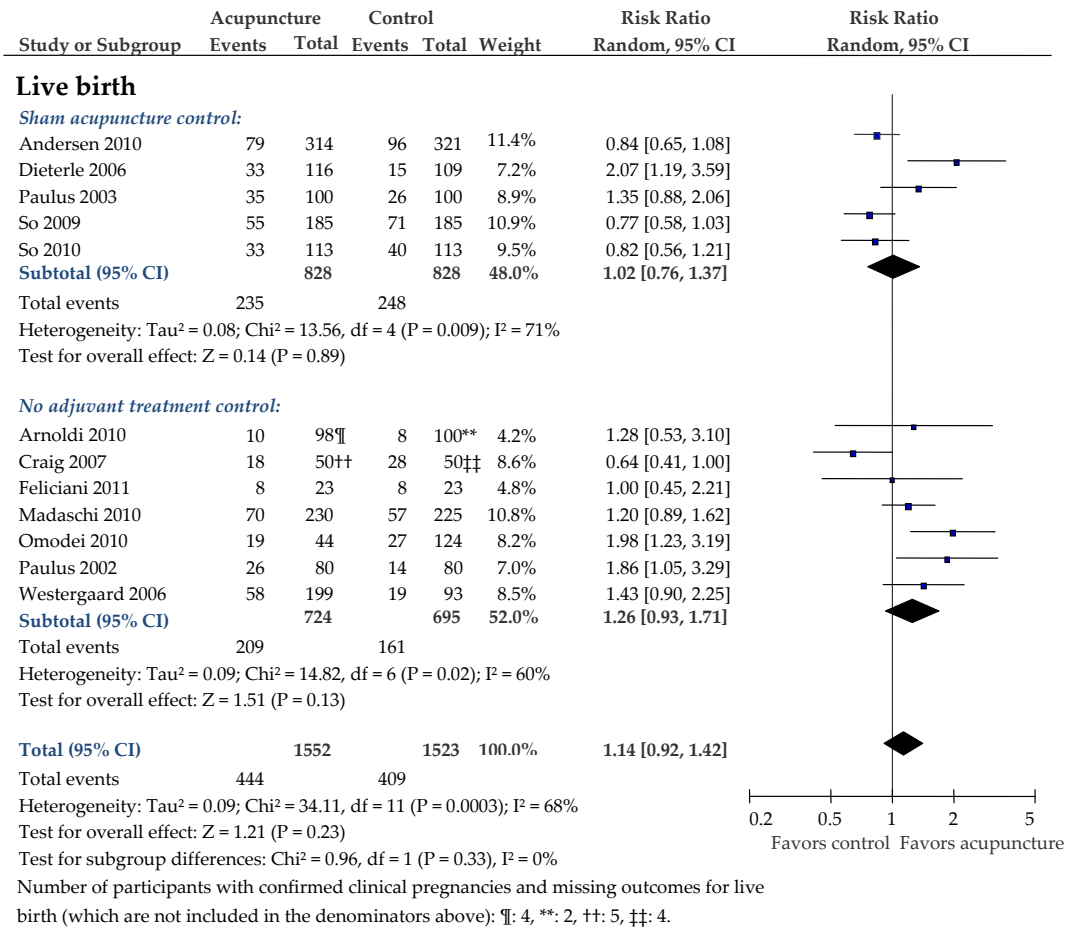


Figure 1. Effects of acupuncture on clinical pregnancy, ongoing pregnancy and live birth outcomes

The centres of the squares represent estimates from individual trials, the centres of the quadrilaterals represent pooled estimates, and the horizontal lines represent 95% confidence intervals.

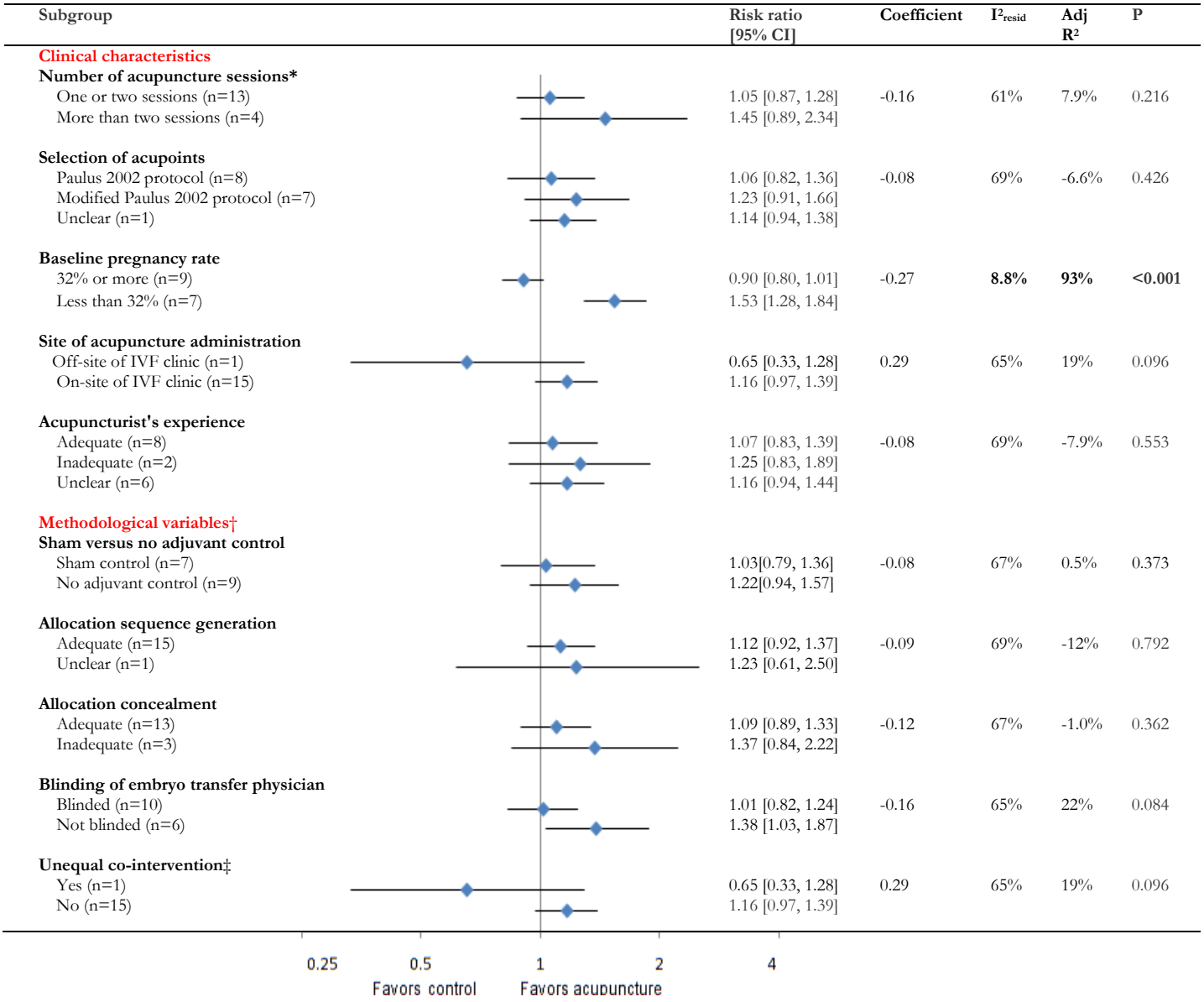


Figure 2. Meta-regression subgroup analyses for the primary outcome (i.e., clinical pregnancy)
The quadrilaterals represent pooled estimates from the trials included in the given subgroup. The estimated regression coefficient from each model was obtained by a weighted least squares meta-regression with risk ratio of pregnancy as the dependent variable, using the modification to the variance of the estimated coefficient suggested by Knapp and Hartung,[56] and supported by Higgins and Thompson.[55] I^2_{resid} values indicate the proportion of the residual variation that is attributable to between-study heterogeneity. The adjusted R^2 values indicate the proportion of between-study variance explained by the covariate. The p values for test of interaction indicate whether the observed differences in results of trials within a given subgroup are compatible with chance alone.

* The So 2010 trial was the only trial that used only one acupuncture treatment session. The Westergaard 2006 trial had two acupuncture treatment arms, one arm received two sessions and the other arm received three, and these arms were grouped separately for this subgroup analysis.

† Incomplete outcome data was also pre-specified as a methodological subgroup variable, but no trial had a high risk of bias due to incomplete outcome data for the clinical pregnancy outcome.

‡ Only the Craig et al trial was judged to have a co-intervention that was applied unequally across treatment group. Namely, in this trial, which evaluated off-site acupuncture, patients in the acupuncture group only were required to drive to and from the off-site acupuncturist's office both before and after the embryo transfer procedure.

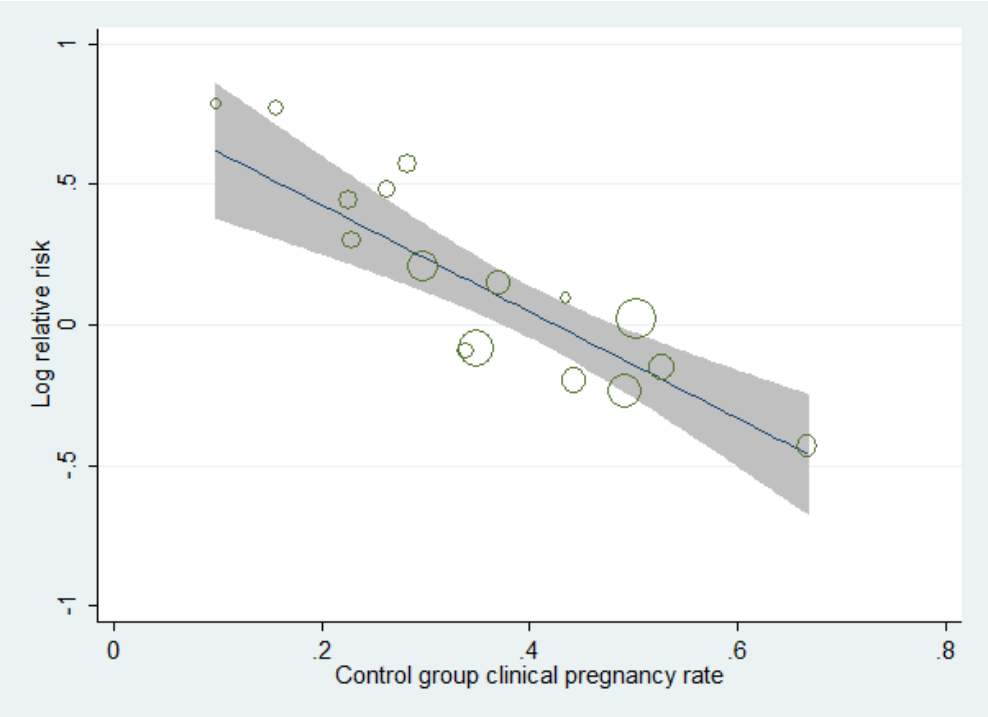


Figure 3. Meta-regression with baseline clinical pregnancy rate as single continuous covariate
Outcome was risk ratio of clinical pregnancy. Circle size represents weight each study was given in the meta-regression analysis. For this meta-regression, the adjusted R^2=90% and the I^2_{resid}=3.3%.

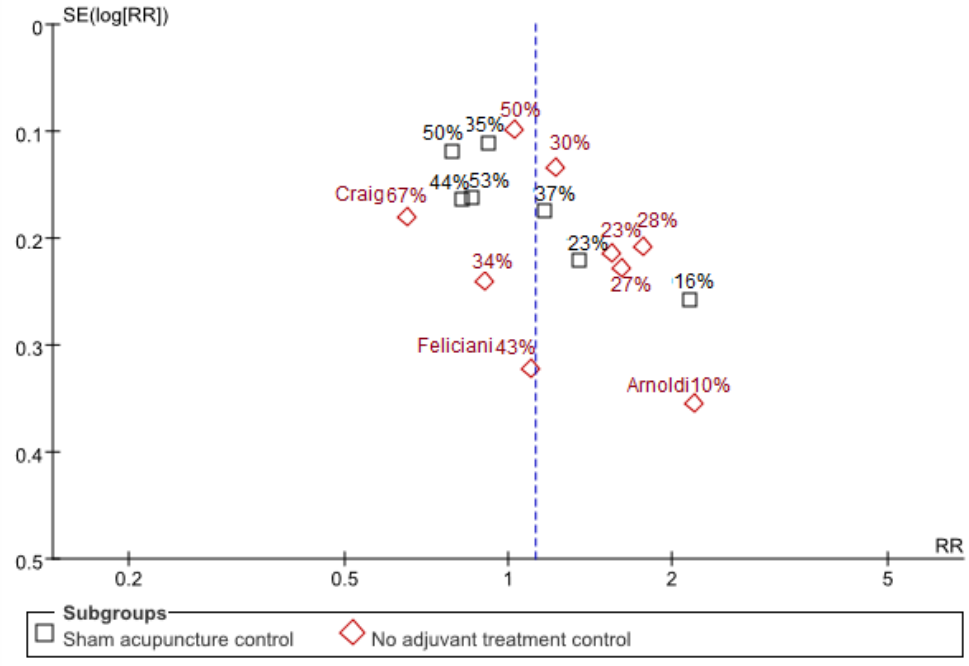


Figure 4. Funnel plot of trials meeting inclusion criteria
The intervention effect estimated from individual trials is plotted on the horizontal scale and a measure of the standard error of the intervention effect is plotted on the vertical axis. The control group success rate is included for each trial on the plot. The Egger test p-value for funnel plot symmetry is 0.032.

Safety of acupuncture

Figure 5 shows the miscarriage rates calculated as the number of losses between clinical pregnancy and live birth divided by the number of clinical pregnancies. Similar results were obtained when calculating miscarriage rates as losses between clinical pregnancy and ongoing pregnancy (across all trials, RR=1.01; 95% CI, 0.76 to 1.35; I^2=0%; 11 trials). A total of 7 trials[59,62-65,68,71] included descriptions of other adverse events, and none of these trials reported any serious adverse events associated with acupuncture.

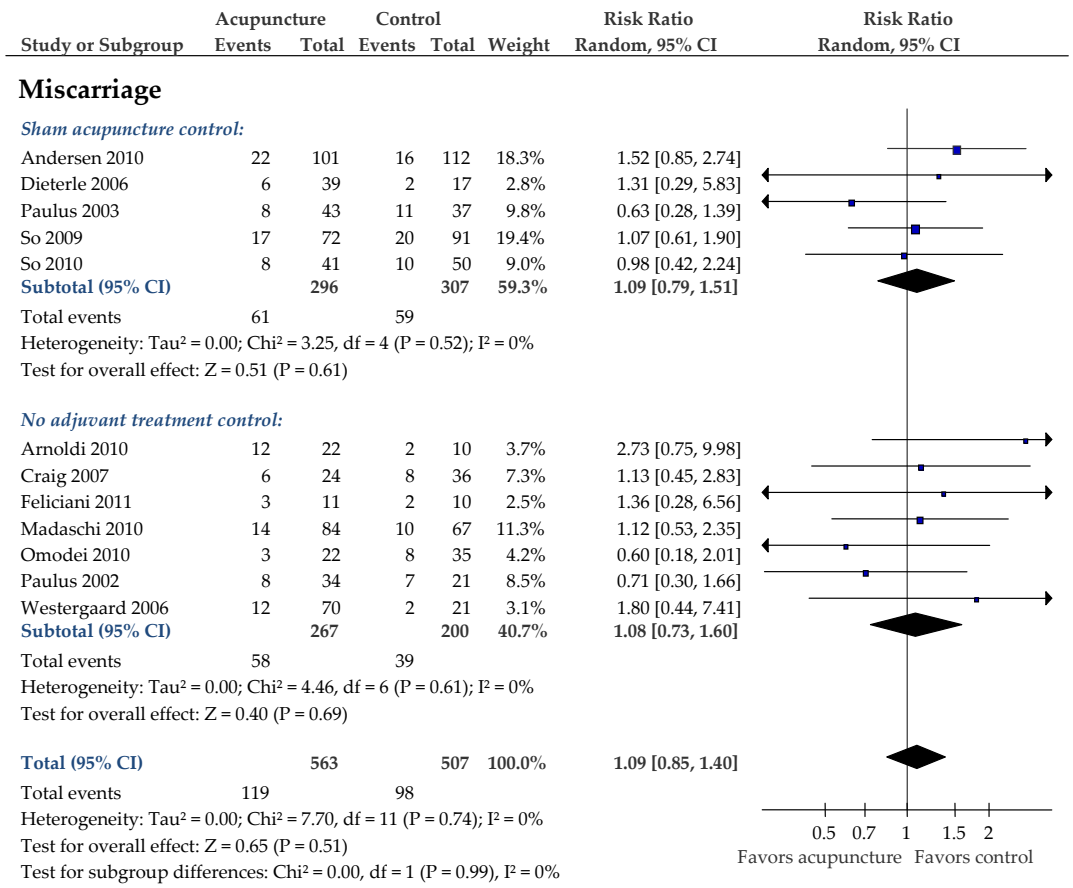


Figure 5. Effects of acupuncture on miscarriage outcome

The centres of the squares represent estimates from individual trials, the centres of the quadrilaterals represent pooled estimates, and the horizontal lines represent 95% confidence intervals.

DISCUSSION

Summary of main results

We found no statistically significant pooled benefits of one to three sessions of acupuncture adjuvant to IVF, across all trials, or when restricting to sham-controlled or no adjuvant treatment-controlled trials. However, there was substantial[48] heterogeneity. Of the 11 variables pre-specified for subgroup analyses, only the baseline pregnancy rate variable showed a statistically significant subgroup effect and appeared to explain most of the heterogeneity in results for the primary outcome measure (clinical pregnancy).

Subgroup analyses based on clinical characteristics

Baseline pregnancy rate was the only statistically significant subgroup finding in our previous review on this topic[21] and has been confirmed in this review update. The magnitude of this subgroup effect is large and there is a low likelihood that chance explains this apparent subgroup effect. In addition, this subgroup effect does not appear to be due to an association between baseline pregnancy rate and either blinding- or randomization-related covariates, or any of the other pre-specified covariates.

The reasons for differences in baseline (i.e., control group) pregnancy rates across trials are complex and driven by several factors, although these differences may be largely explained by differences in IVF regulations across countries, in particular regarding the number of embryos transferred per cycle.[4,5,75,76] That is, European countries have lower IVF pregnancy rates than does the United States, and a major reason for this difference is that fewer embryos are transferred per cycle in European countries than in the United States,[5] in an effort to reduce twins and triplets.[4] In Italy and Germany, the country setting of six of the eight included European RCTs, the average IVF pregnancy rates are lower than in other European countries because of Italian/German laws which set a limit of using only three ooctyes per IVF cycle, all of which must be transferred, regardless of whether these oocytes develop into high quality embryos.[5,6,75,76] In the United States, on the other hand, embryo selection[77] is routine, and the number of embryos that can be transferred is less strict and less well-regulated.[4] Another reason for higher baseline pregnancy rates in the United States is that US patients with poor ovarian reserve (i.e., those who produce fewer eggs and end up with fewer embryos to transfer or chose from) are more likely to be diverted from undergoing IVF.[4] Differences in baseline pregnancy rates can also be due to differences in inclusion criteria across trials. For example, one Italian trial in this review[67] restricted trial inclusion to women with an unfavourable reproductive prognosis, which might explain why the baseline pregnancy rates in this Italian trial were lower than Italian national averages.[3] Two German trials[52,72] restricted trial inclusion to only women with high quality embryos available to transfer, which might explain why the baseline pregnancy rates in these German trials were higher than German national averages.[3] Finally, differences in baseline pregnancy rates across included trials can also be due to participant selection in the individual trials, in terms of participants' predictors for success in IVF.[78] Namely, for two of the trials with lower baseline pregnancy rates,[59,63] the trial authors reported that the trials' baseline pregnancy rates were lower than the average pregnancy rates at the trials' clinics because of the higher mean age of the participants included in the trials relative to the mean age of patients seen at the clinic site. And for one of these two trials,[63] the trial authors also reported an overrepresentation by women with poor pregnancy outcomes from previous cycles. Thus, there are multiple factors that can explain differences across trials in baseline pregnancy rates. One potential explanation for the differential effect of acupuncture in trials with higher versus lower baseline rates may be that in IVF settings where the baseline pregnancy rates are already high, the relative added value of additional co-interventions, such as acupuncture, may be lower.[79]

Although it was not a statistically significant subgroup finding, there was a trend suggesting that acupuncture administered on- site of the IVF clinic had more positive effects than acupuncture administered off-site of the IVF clinic. However, inferences regarding this observed differential effect are weak because only a single trial[68] contributes to the off-site subgroup. The qualitatively and quantitatively different results in the off-site Craig et al trial, which had very high baseline pregnancy rates (i.e., 67%), could have been driven by another variable, other than the off-site acupuncture administration. One potential biological rationale for the negative results of the Craig et al trial, postulated by the trial's principal investigator,[80] is that the co-intervention, in only the acupuncture group, of driving to and from the off-site acupuncturist's office, both before and after the embryo transfer, rather than resting in bed, may have increased patients' stress, which may have affected pregnancy success rates negatively through a psychobiological effect mechanism.[34,81,82] However, studies of the effects of stress on IVF success rates have had inconsistent results, and the influence of psychological stress on IVF success appears limited, at best.[82] In addition, while resting in bed for two or three days after the embryo transfer procedure has been recommended by some IVF physicians on the theory that bed rest can prevent the embryo from being expelled from the uterus, there is no evidence to support this recommendation.[83] Thus, the support from a biological rationale for this putative subgroup effect regarding on- versus off-site acupuncture is unconvincing. However, because only a single trial has evaluated off-site acupuncture, and because Craig et al conducted this trial specifically to address whether the positive findings from the earlier Paulus et al trials[52,72] which involved on-site acupuncture could be confirmed in "real world" off-site settings,[80] and because this Craig et al trial found a *reduction* in pregnancy rates with off-site adjuvant acupuncture compared with no IVF adjuvant treatment, IVF physicians and patients may wish to consider this trial's findings in deciding whether IVF patients should have off-site acupuncture on the day of embryo transfer.

Subgroup analyses based on risk of bias

Seven of the trials[58,59,62-65,72] used a sham control to blind participants. However, the necessity to blind participants is arguable when the outcomes are entirely objective (that is, pregnancy and birth).[21,84-86] Indeed, the Cochrane Menstrual Disorders and Subfertility Group website states that "a study may not be blinded but if follow-up is complete and outcomes are unequivocal the lack of blinding may be assessed not to increase risk of bias."[43] As previously argued,[21,22,87] it seems unlikely that a woman's knowledge of whether or not she was receiving acupuncture would affect her ability to become pregnant. Even if adjuvant acupuncture were to increase IVF success rates strictly through a non-needling-related psychobiological placebo effect mechanism (i.e., women who believe they are receiving acupuncture have reduced stress, which thereby subtly influences their pregnancy-related hormone production), this effect would be integral to the working mechanism by which adjuvant acupuncture increases IVF pregnancy success rates. Therefore, it seems inappropriate to control for and equalize any such stress-reduction placebo effect by using a sham control to blind trial participants, unless the purpose of the trial is to study the mechanism responsible for adjuvant acupuncture's effects on IVF success rates.

The risk of using a sham control is that some sham interventions may influence the pregnancy outcome through the same putative mechanism of true acupuncture (e.g., by needle insertion, stimulation of true acupuncture points).[87] For example, the sham intervention used in four of the sham-controlled trials[58,64,65,72] was the non-penetrating Streitberger sham needles placed at the true acupuncture points. The two validation trials[73,74] of this Streitberger sham needle found that these Streitberger sham needles elicit the de qi needling sensation in a large proportion of patients, suggesting that these sham needles may have acupuncture-specific needling effects, particularly if the Streitberger sham needles are manipulated in the same way as the true acupuncture needles, as

was the case in two of the included sham-controlled trials.[64,65] We initially attempted to assess whether or not the sham control intervention used in each trial was likely to have a risk of affecting the pregnancy outcome through the same putative mechanism of true acupuncture; however, because no consensus could be reached in making these assessments (see Supplementary Appendix Table 5), the assessments could not be used in the analyses. Instead, our subgroup analysis test involved grouping all sham-controlled trials together in a single subgroup; however, some sham control interventions may be physiologically active, and others physiologically inert, in terms of their effects on the pregnancy outcome, and therefore this subgroup test (i.e., no adjuvant treatment versus any type of sham control) may not detect an important effect of control type. Finally, although this review found no statistically significant subgroup effect for type of control, the no adjuvant treatment-controlled subgroup showed a slightly larger pooled effect than the sham-controlled subgroup, particularly if the outlying Craig et al trial was removed (RR 1.31 (1.08 to 1.58);I^2=49%;8 trials). Therefore, the possibility that some of the sham interventions may have had acupuncture-specific effects cannot be ruled out.

Although it was not a statistically significant subgroup finding, there was a trend suggesting that IVF physicians who are not blinded may perform more successful embryo transfers when adjuvant acupuncture is used, at least in a trial setting. This subgroup finding was unexpected because, considering the cost of embryo transfer and the importance of successful transfers to maintaining high pregnancy success rates at clinics, we *a priori* expected that physicians would be motivated primarily to perform a successful embryo transfer for all trial participants, rather than to show that acupuncture, a non-proprietary treatment, is an effective adjuvant procedure. Indeed, in IVF trials, where outcomes are objective, blinding of either patients or physicians is "infrequently attempted,"[37] and such blinding components are often not considered as critical elements related to the evaluation of risk of bias.[50,88] Yet, this borderline significant subgroup finding suggests that physician blinding may decrease bias, and therefore this variable may possibly be appropriate for inclusion in scales that assess risk of bias of IVF trials.[43] However, blinding the embryo transfer physician still does not require the use of a sham acupuncture control[52,60,61,68] because the adjuvant acupuncture is not administered at the same time as the embryo transfer procedure.

Interpretation of small study effects

The funnel plot indicated that estimates of the intervention effect were more beneficial in smaller studies, which may be due to smaller studies with statistically significant, positive results being more likely to be published (i.e., publication bias). Funnel plot asymmetry can also be due to other types of reporting bias besides publication bias (e.g., language bias, multiple publication bias, publication of negative results only as conference abstracts). However, in this review, these other types of reporting bias seem less likely than publication bias, given our extensive searches (including Chinese databases and conference proceedings), and our contacts with the authors of included trials, which confirmed that none of these trials are duplicate publications. Although we attempted to locate any unpublished trials by contacting experts in the field, none were located, and the possibility of unidentified, unpublished studies with negative results is one possible explanation for the asymmetric funnel plot. However, the Cochrane Collaboration's guidelines for addressing reporting biases and interpreting funnel plot asymmetry stress that "an asymmetric funnel plot should not be equated with publication bias", and that "publication bias should be considered as only one of a number of possible explanations."[51] Also according to Cochrane guidelines, "Funnel plot asymmetry...may also result from clinical heterogeneity between studies (for example different control event rates)...".[89] We therefore added the control event rates to the funnel plot, and the resulting plot shows that the trials with the lowest control group success rates show the most positive results (Figure 4). Therefore, it is impossible to know whether the funnel plot asymmetry is

due to publication bias, or whether this pattern is due to greater benefits in the trials with lower control group success rates.[89]

We also tried to identify other potential explanations for the funnel plot asymmetry. Namely, we prepared separate funnel plots (and conducted corresponding separate Egger tests) for the sham-controlled trial subgroup and the no adjuvant treatment-controlled trial subgroup. These separate tests found that the funnel plot asymmetry seemed to be driven largely by the sham-controlled trials (Egger p=0.028 for sham-controlled trials and p=0.303 for no adjuvant treatment controlled trials). Cochrane guidelines suggest that funnel plot asymmetry may be due to larger 'negative' trials (conducted later to confirm the earlier 'positive' trials' results) using a more effective control treatment or a less thoroughly implemented test intervention.[51] Therefore, this issue was examined among the sham-controlled trials that contributed to the funnel plot asymmetry. Namely, three of the largest sham-controlled trials,[58,64,65] which all had 'negative' results, used non-penetrating Streitberger sham needles placed at the true acupuncture points as the control intervention. For two of these trials,[64,65] these sham needles were manipulated in the same way as the true acupuncture needles. The author of these two trials discussed that this sham control intervention used was likely an effective treatment due to its acupuncture-specific effects, and a conclusion of their 2009 trial was that "Placebo acupuncture may not be inert".[64] For the third and largest trial,[58] for which the acupuncture was performed by nurse-acupuncturists and other nurses trained by these nurse-acupuncturists, the true acupuncture needles were not manipulated, in order to avoid unblinding the patients to treatment assignment. Therefore, in these three recent, large sham-controlled trials with 'negative' results, manipulating the needles in the sham acupuncture group[64,65] and/or not manipulating the needles in the true acupuncture group[58] may have caused the control treatment to be an effective treatment and/or the test intervention to be less effective. This possibility is particularly relevant in light of a recent trial which found that manipulating the acupuncture needles to obtain the 'de qi' sensation resulted in a much greater therapeutic effect compared to not manipulating the needles (odds ratio=4.16, for primary outcome).[90] The use of a potentially more effective control and/or less thoroughly implemented test intervention in these three large sham-controlled trials with 'negative' results is only one of many possible explanations for the funnel plot asymmetry among the sham-controlled trials. This example is provided primarily to illustrate the challenges in interpreting funnel plot asymmetry in meta-analyses of complex interventions that used heterogeneous treatments and controls, when the funnel plot test was devised on the premise that the studies come from a single underlying population.[51,91]

Comparison with other studies or reviews

Although there have been multiple systematic reviews of acupuncture as an adjuvant to IVF,[21,24-31] only the 2012 reviews by Zheng and colleagues[31] and Qu and colleagues[29] included some of the more recent trials. And therefore, comparisons with the Zheng et al review and Qu et al review are most relevant. The Zheng et al review concluded that adjuvant acupuncture improves clinical pregnancy rates and live birth rates, but only after excluding the Streitberger sham needle-controlled trials. Zheng et al excluded these trials in drawing their conclusions because they contended that the Streitberger sham intervention is not an inert control; however, Zheng et al did not provide evidence to support this contention. Also, because the Zheng et al review was not based on a publicly available protocol, it cannot be determined whether the decision to exclude the trials using the Streitberger shams was pre-specified, or whether instead this exclusion was driven by the Streitberger sham-controlled trials' results. In addition, the Zheng et al review, as well as all other previous systematic reviews, were based on only the published data, with the exception of one previous review[27] for which two RCT authors[60,68] were contacted and supplied additional information. We have successfully obtained unpublished data on methods and/or outcomes from

authors of all eligible RCTs, which allowed for the inclusion of more RCTs and more complete information from these included RCTs. For example, Zheng et al excluded two methodologically sound RCTs[62,71] because the "data for the exact pregnancy events and totals were not available" from the publication; however, we were able to obtain precise outcome data for these two RCTs by contacting the authors. Finally, Zheng et al included both the early results from one trial,[92] as well as the final results,[70] thus double counting some of the same participants from this trial. The other 2012 review, by Qu et al, concluded that no benefits of adjuvant acupuncture were apparent from the effect estimate pooled from all trials. However, the Qu et al review did not include six of the RCTs included in our review,[61,65,67,69-71] four of which[65,67,70,71] were published prior to October 2010, the end date of Qu et al's search window. Also, Qu et al did not examine potential subgroup variables to understand their impact on the effects of adjuvant acupuncture on IVF, on the clinically and methodologically heterogeneous trials reviewed. In contrast, the detailed information we obtained allowed us not only to evaluate potential effect modification due to the pre-specified subgroup variables, but also to examine whether other subgroup variables were confounders of any apparent subgroup effects.

Implications for research

Future studies might further investigate the relationship between baseline rate of pregnancy and the efficacy of adjuvant acupuncture, and also further investigate which variables are responsible for this relationship. Because baseline pregnancy rate is possibly a proxy for the number of embryos transferred, although other factors are also involved,[4] an individual patient data meta-analysis[93] may be helpful to further investigate the relationship between the number of embryos transferred, as well as other individual participant level variables (e.g., age[66]), and the efficacy of adjuvant acupuncture.

For future trials, if the objective is to investigate the mechanism of IVF adjuvant acupuncture, and specifically whether the effect is mediated through the point-specific needle placement, or alternatively through a non-needling related psychobiological placebo mechanism, both a sham control arm and a no adjuvant treatment arm would be necessary. A sham control arm might also be necessary for recruitment, if potential trial participants would be unwilling to be randomized to a no adjuvant treatment control.[94] However, because it is difficult to assess whether or not a sham control intervention has acupuncture specific effects that may increase IVF success rates, sham-controlled trials in this area can also potentially complicate the interpretation of the overall evidence.[87] For addressing the more clinically relevant question of the total effects of acupuncture (i.e., specific needling effects plus any non-needle-related placebo effects) in contributing to any increases in IVF pregnancy success rates, a sham control seems unnecessary.[87] This question about the need for sham controls may also apply to other invasive, difficult to blind adjuvant procedures evaluated in IVF RCTs.[95]

REFERENCES

1. Evers JL. Female subfertility. *Lancet* 2002;360(9327):151-9.
2. ESHRE Capri Workshop Group. Intracytoplasmic sperm injection (ICSI) in 2006: evidence and evolution. *Hum Reprod Update* 2007;13(6):515-26.
3. Nygren KG, Sullivan E, Zegers-Hochschild F, Mansour R, Ishihara O, Adamson GD, de Mouzon J. International Committee for Monitoring Assisted Reproductive Technology (ICMART) world report: assisted reproductive technology 2003. *Fertil Steril* 2011;95(7):2209-22, 22 e1-17.
4. Abdalla HI. Are US results for assisted reproduction better than the rest? Is it a question of competence or policies? *Reprod Biomed Online* 2010;21(5):624-30.
5. Maheshwari A, Griffiths S, Bhattacharya S. Global variations in the uptake of single embryo transfer. *Hum Reprod Update* 2011;17(1):107-20.
6. Gerris JM. Single embryo transfer and IVF/ICSI outcome: a balanced appraisal. *Hum Reprod Update* 2005;11(2):105-21.
7. Pinborg A. IVF/ICSI twin pregnancies: risks and prevention. *Hum Reprod Update* 2005;11(6):575-93.
8. Centers for Disease Control and Prevention. Assisted reproductive technology success rates: national summary and fertility clinic reports 2009. Atlanta, GA: US Department of Health and Human Services. Centers for Disease Control and Prevention, 2011.
9. Harper J, Magli MC, Lundin K, Barratt CL, Brison D. When and how should new technology be introduced into the IVF laboratory? *Hum Reprod* 2012;27(2):303-13.
10. Smith JF, Eisenberg ML, Millstein SG, Nachtigall RD, Shindel AW, Wing H, Cedars M, Pasch L, Katz PP. The use of complementary and alternative fertility treatment in couples seeking fertility care: data from a prospective cohort in the United States. *Fertil Steril* 2010;93(7):2169-74.
11. Domar AD, Conboy L, Denardo-Roney J, Rooney KL. Lifestyle behaviors in women undergoing in vitro fertilization: a prospective study. *Fertil Steril* 2012;97(3):697-701.
12. Hopton AK, Curnoe S, Kanaan M, Macpherson H. Acupuncture in practice: mapping the providers, the patients and the settings in a national cross-sectional survey. *BMJ Open* 2012;2(1):e000456.
13. Witt CM, Lao L, MacPherson H. Evidence on acupuncture safety needs to be based on large-scale prospective surveys, not single case reports. *Pain* 2011;152(9):2180; auther reply 84-6.
14. Elden H, Ladfors L, Olsen MF, Ostgaard HC, Hagberg H. Effects of acupuncture and stabilising exercises as adjunct to standard treatment in pregnant women with pelvic girdle pain: randomised single blind controlled trial. *BMJ* 2005;330(7494):761.
15. Kvorning N, Holmberg C, Grennert L, Aberg A, Akeson J. Acupuncture relieves pelvic and low-back pain in late pregnancy. *Acta Obstet Gynecol Scand* 2004;83(3):246-50.
16. Smith C, Crowther C, Beilby J. Pregnancy outcome following women's participation in a randomised controlled trial of acupuncture to treat nausea and vomiting in early pregnancy. *Complement Ther Med* 2002;10(2):78-83.
17. de Lacey S, Smith CA, Paterson C. Building resilience: a preliminary exploration of women's perceptions of the use of acupuncture as an adjunct to In Vitro Fertilisation. *BMC Complement Altern Med* 2009;9:50.
18. Isoyama D, Cordts EB, de Souza van Niewegen AM, de Almeida Pereira de Carvalho W, Matsumura ST, Barbosa CP. Effect of acupuncture on symptoms of anxiety in women undergoing in vitro fertilisation: a prospective randomised controlled study. *Acupunct Med* 2012;30(2):85-8.
19. Kovarova P, Smith CA, Turnbull DA. An exploratory study of the effect of acupuncture on self-efficacy for women seeking fertility support. *Explore (NY)* 2010;6(5):330-4.
20. Smith CA, Ussher JM, Perz J, Carmady B, de Lacey S. The effect of acupuncture on psychosocial outcomes for women experiencing infertility: a pilot randomized controlled trial. *J Altern Complement Med* 2011;17(10):923-30.
21. Manheimer E, Zhang G, Udoff L, Haramati A, Langenberg P, Berman BM, Bouter LM. Effects of acupuncture on rates of pregnancy and live birth among women undergoing in vitro fertilisation: systematic review and meta-analysis. *BMJ* 2008;336(7643):545-9.
22. Manheimer E, van der Windt DA, Stafford K, Liu J, Tierney J, Lao L, Berman B, Bouter L. Effects of adjuvant acupuncture on pregnancy rates among women undergoing in vitro fertilization: systematic review and meta-analysis. PROSPERO 2011:CRD42011001309 Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001309 (accessed 2013 July 12).
23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6(7):e1000097.
24. Cheong Y, Nardo LG, Rutherford T, Ledger W. Acupuncture and herbal medicine in in vitro fertilisation: a review of the evidence for clinical practice. *Hum Fertil (Camb)* 2010;13(1):3-12.
25. Cheong YC, Hung Yu Ng E, Ledger WL. Acupuncture and assisted conception. *Cochrane Database Syst Rev* 2008(4):CD006920.
26. El-Toukhy T, Khalaf Y. The impact of acupuncture on assisted reproductive technology outcome. *Curr Opin Obstet Gynecol* 2009;21(3):240-6.

27. El-Toukhy T, Sunkara SK, Khairy M, Dyer R, Khalaf Y, Coomarasamy A. A systematic review and meta-analysis of acupuncture in in vitro fertilisation. *BJOG* 2008;115(10):1203-13.
28. Ng EH, So WS, Gao J, Wong YY, Ho PC. The role of acupuncture in the management of subfertility. *Fertil Steril* 2008;90(1):1-13.
29. Qu F, Zhou J, Ren RX. Effects of acupuncture on the outcomes of in vitro fertilization: a systematic review and meta-analysis. *J Altern Complement Med* 2012;18(5):429-39.
30. Sunkara SK, Coomarasamy A, Khalaf Y, El-Toukhy T. Acupuncture and in vitro fertilization: updated meta-analysis. *Hum Reprod* 2009;24(8):2047-8.
31. Zheng CH, Huang GY, Zhang MM, Wang W. Effects of acupuncture on pregnancy rates in women undergoing in vitro fertilization: a systematic review and meta-analysis. *Fertil Steril* 2012;97(3):599-611.
32. Zheng CH, Zhang MM, Huang GY, Wang W. The Role of Acupuncture in Assisted Reproductive Technology. *Evid Based Complement Alternat Med* 2012;2012:543924.
33. Ulett GA, Han S, Han JS. Electroacupuncture: mechanisms and clinical application. *Biol Psychiatry* 1998;44(2):129-38.
34. Stener-Victorin E, Waldenstrom U, Andersson SA, Wikland M. Reduction of blood flow impedance in the uterine arteries of infertile women with electro-acupuncture. *Hum Reprod* 1996;11(6):1314-7.
35. Kwan I, Bhattacharya S, Knox F, McNeil A. Pain relief for women undergoing oocyte retrieval for assisted reproduction. *Cochrane Database Syst Rev* 2013;1:CD004829.
36. Stener-Victorin E. The pain-relieving effect of electro-acupuncture and conventional medical analgesic methods during oocyte retrieval: a systematic review of randomized controlled trials. *Hum Reprod* 2005;20(2):339-49.
37. Arce JC, Nyboe Andersen A, Collins J. Resolving methodological and clinical issues in the design of efficacy trials in assisted reproductive technologies: a mini-review. *Hum Reprod* 2005;20(7):1757-71.
38. Daya S. Pitfalls in the design and analysis of efficacy trials in subfertility. *Hum Reprod* 2003;18(5):1005-9.
39. Griesinger G, Dafopoulos K, Schultze-Mosgau A, Felberbaum R, Diedrich K. What is the most relevant standard of success in assisted reproduction? Is BESST (birth emphasizing a successful singleton at term) truly the best? *Hum Reprod* 2004;19(6):1239-41.
40. Zegers-Hochschild F, Nygren KG, Adamson GD, de Mouzon J, Lancaster P, Mansour R, Sullivan E. The ICMART glossary on ART terminology. *Hum Reprod* 2006;21(8):1968-70.
41. Higgins JPT and Altman DG (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (accessed 30 December 2012).
42. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine* 2003;28(12):1290-9.
43. Cochrane Menstrual Disorders and Subfertility Group. Resources for review authors. 2010. The Cochrane Collaboration, Auckland. Available from http://mdsg.cochrane.org/resources-review-authors/ (accessed 30 December 2012).
44. Clarke JF, van Rumste MM, Farquhar CM, Johnson NP, Mol BW, Herbison P. Measuring outcomes in fertility trials: can we rely on clinical pregnancy rates? *Fertil Steril* 2010;94(5):1647-51.
45. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
46. Deeks JJ and Higgins JPT. Statistical algorithms in Review Manager 5. August 2010. Available from http://www.cochrane.org/sites/default/files/uploads/handbook/Statistical_Methods_in_RevMan5-1.pdf.
47. The Nordic Cochrane Centre. The Cochrane Collaboration. Review Manager (RevMan) [Computer program]. Version 5.1. 2011. The Cochrane Collaboration, Copenhagen: Denmark.
48. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (accessed 30 December 2012).
49. Deeks JJ, Higgins JPT, Altman DG, editors. Chapter 8: Analysing and presenting results (section 8.11.3.1). In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 4.2.6 [updated September 2006]. The Cochrane Collaboration, 2006. Available from http://www.cochrane.org/sites/default/files/uploads/Handbook4.2.6Sep2006.pdf (accessed 30 December 2012).
50. Dias S, McNamee R, Vail A. Evidence of improving quality of reporting of randomized controlled trials in subfertility. *Hum Reprod* 2006;21(10):2617-27.
51. Sterne JAC, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (accessed 30 December 2012).
52. Paulus WE, Zhang M, Strehler E, El-Danasouri I, Sterzik K. Influence of acupuncture on the pregnancy rate in patients who undergo assisted reproduction therapy. *Fertil Steril* 2002;77(4):721-4.

53. de Mouzon J, Goossens V, Bhattacharya S, Castilla JA, Ferraretti AP, Korsak V, Kupka M, Nygren KG, Andersen AN. Assisted reproductive technology in Europe, 2007: results generated from European registers by ESHRE. *Hum Reprod* 2012;27(4):954-66.
54. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21(11):1559-73.
55. Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23(11):1663-82.
56. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 2003;22(17):2693-710.
57. Harbord RM, Higgins JP. Meta-regression in Stata. *The Stata Journal* 2008;8(4):493-519.
58. Andersen D, Lossl K, Nyboe Andersen A, Furbringer J, Bach H, Simonsen J, Larsen EC. Acupuncture on the day of embryo transfer: a randomized controlled trial of 635 patients. *Reprod Biomed Online* 2010;21(3):366-72.
59. Dieterle S, Ying G, Hatzmann W, Neuer A. Effect of acupuncture on the outcome of in vitro fertilization and intracytoplasmic sperm injection: a randomized, prospective, controlled clinical study. *Fertil Steril* 2006;85(5):1347-51.
60. Domar AD, Meshay I, Kelliher J, Alper M, Powers RD. The impact of acupuncture on in vitro fertilization outcome. *Fertil Steril* 2009;91(3):723-6.
61. Madaschi C, Braga DP, Figueira Rde C, Iaconelli AJ, Borges EJ. Effect of acupuncture on assisted reproduction treatment outcomes. *Acupunct Med* 2010;28(4):180-4.
62. Moy I, Milad MP, Barnes R, Confino E, Kazer RR, Zhang X. Randomized controlled trial: effects of acupuncture on pregnancy rates in women undergoing in vitro fertilization. *Fertil Steril* 2011;95(2):583-7.
63. Smith C, Coyle M, Norman RJ. Influence of acupuncture stimulation on pregnancy rates for women undergoing embryo transfer. *Fertil Steril* 2006;85(5):1352-8.
64. So EW, Ng EH, Wong YY, Lau EY, Yeung WS, Ho PC. A randomized double blind comparison of real and placebo acupuncture in IVF treatment. *Hum Reprod* 2009;24(2):341-8.
65. So EW, Ng EH, Wong YY, Yeung WS, Ho PC. Acupuncture for frozen-thawed embryo transfer cycles: a double-blind randomized controlled trial. *Reprod Biomed Online* 2010;20(6):814-21.
66. Westergaard LG, Mao Q, Krogslund M, Sandrini S, Lenz S, Grinsted J. Acupuncture on the day of embryo transfer significantly improves the reproductive outcome in infertile women: a prospective, randomized trial. *Fertil Steril* 2006;85(5):1341-6.
67. Arnoldi M, Diaferia D, Corbucci MG, De Lauretis L. The role of acupuncture in patients at unfavourable reproductive prognosis in IVF: a prospective randomised study. *Hum Reprod* 2010;25 (Suppl 1):i257 (Abstract P-363).
68. Craig LB, Criniti AR, Hansen KR, Marshall LA, Soules MR. Acupuncture lowers pregnancy rates when performed before and after embryo transfer. *Fertil Steril* 2007;88(Suppl 1):S-40 (Abstract O-106).
69. Feliciani E, Ferraretti AP, Paesano C, Pellizzaro E, Magli MC, Gianaroli L. The role of acupuncture in ART: preliminary results of an ongoing prospective randomized study. *Hum Reprod* 2011;26(Suppl 1):i230 (Abstract P-81).
70. Fratterelli JL, Leondires MR, Fong K, Theall A, Locatelli S, Scott RT. Laser acupuncture before and after embryo transfer improves ART delivery rates: Results of a prospective randomized double-blinded placebo controlled five-armed trial involving 1000 patients. *Fertil Steril* 2008;90(Suppl 1):S-105 (Abstract O-285).
71. Omodei U, Piccioni G, Tombesi S, Dordoni D, Fallo L, Ghilardi F. Effect of acupuncture on rates of pregnancy among women undergoing in vitro fertilization. *Fertil Steril* 2010;94(4 Suppl):S170 (Abstract P-266).
72. Paulus WE, Zhang M, Strehler E, Seybold B, Sterzik K. Placebo-controlled trial of acupuncture effects in assisted reproduction therapy. *Hum Reprod* 2003;18(6 Suppl 1):xviii18-19.
73. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet* 1998;352(9125):364-5.
74. White P, Lewith G, Hopwood V, Prescott P. The placebo needle, is it a valid and convincing placebo for use in acupuncture trials? A randomised, single-blind, cross-over pilot trial. *Pain* 2003;106(3):401-9.
75. Dal Canto MB, Mignini Renzini M, Brambillasca F, Cepparo H, Comi R, Villa A, Rangoni G, Mastrolilli M, Crippa M, de Ponti E, et al. IVM--the first choice for IVF in Italy. *Reprod Biomed Online* 2006;13(2):159-65.
76. Ludwig M, Schopper B, Katalinic A, Sturm R, Al-Hasani S, Diedrich K. Experience with the elective transfer of two embryos under the conditions of the german embryo protection law: results of a retrospective data analysis of 2573 transfer cycles. *Hum Reprod* 2000;15(2):319-24.
77. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod* 2011;26(6):1270-83.
78. van Loendersloot LL, van Wely M, Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;16(6):577-89.
79. Sherman KJ, Cherkin DC, Ichikawa L, Avins AL, Barlow WE, Khalsa PS, Deyo RA. Characteristics of patients with chronic back pain who benefit from acupuncture. *BMC Musculoskelet Disord* 2009;10:114.

80. Sullivan MG. Study questions benefit of acupuncture in IVF. *Ob.Gyn.News.* 2007 November 1, pg. 21. (Available at: http://obgyn.imng.com/fileadmin/content_pdf/obn/archive_pdf/vol42iss21/70921_main.pdf) (accessed 30 December 2012).
81. Ferin M, Vande Wiele R. Endogenous opioid peptides and the control of the menstrual cycle. *Eur J Obstet Gynecol Reprod Biol* 1984;18(5-6):365-73.
82. Matthiesen SM, Frederiksen Y, Ingerslev HJ, Zachariae R. Stress, distress and outcome of assisted reproductive technology (ART): a meta-analysis. *Hum Reprod* 2011;26(10):2763-76.
83. Li B, Zhou H, Li W. Bed rest after embryo transfer. *Eur J Obstet Gynecol Reprod Biol* 2011;155(2):125-8.
84. Hrobjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev* 2004(3):CD003974.
85. Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, Stone DA, Nam BH, Kirsch I, Goldman RH. Sham device v inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006;332(7538):391-7.
86. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336(7644):601-5.
87. Manheimer E. Selecting a control for in vitro fertilization and acupuncture randomized controlled trials (RCTs): how sham controls may unnecessarily complicate the RCT evidence base. *Fertil Steril* 2011;95(8):2456-61.
88. Vail A, Gardener E. Common statistical errors in the design and analysis of subfertility trials. *Hum Reprod* 2003;18(5):1000-4.
89. Alderson P and Green S, editors. Cochrane Collaboration's open learning material for Cochrane reviewers. 2002: Module 15. Publication bias. Available from http://130.226.106.152/openlearning/HTML/mod15.htm.
90. Xu SB, Huang B, Zhang CY, Du P, Yuan Q, Bi GJ, Zhang GB, Xie MJ, Luo X, Huang GY, et al. Effectiveness of strengthened stimulation during acupuncture for the treatment of Bell palsy: a randomized controlled trial. *CMAJ* 2013;185(6):473-9.
91. Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003;22(13):2113-26.
92. Benson MR, Elkind-Hirsch KE, Theall A, Fong K, Hogan RB, Scott RT. Impact of acupuncture before and after embryo transfer on the outcome of in vitro fertilization cycles: a prospective single-blind randomized study. *Fertil Steril* 2006;86(3 Suppl 1):S135 (Abstract P-18).
93. Broeze KA, Opmeer BC, van der Veen F, Bossuyt PM, Bhattacharya S, Mol BW. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. *Hum Reprod Update* 2010;16(6):561-7.
94. Smith CA, de Lacey S, Chapman M, Ratcliffe J, Norman RJ, Johnson N, Sacks G, Lyttleton J, Boothroyd C. Acupuncture to improve live birth rates for women undergoing in vitro fertilization: a protocol for a randomized controlled trial. *Trials* 2012;13(1):60.
95. Holt J, Lord J, Acharya U, White A, O'Neill N, Shaw S, Barton A. The effectiveness of foot reflexology in inducing ovulation: a sham-controlled randomized trial. *Fertil Steril* 2009;91(6):2514-9.

Chapter 10. General Discussion

The purpose of this General Discussion Chapter is to reflect on the findings and to discuss the methodological issues that were identified while working on this thesis, and to set the agenda for future research. This Chapter concludes with a brief discussion of the vexing challenge that policy-makers and guideline developers face when confronted with the difficult to interpret RCTs and systematic reviews of acupuncture, which often find acupuncture superior to standard care but equivalent to sham acupuncture.

Taxonomy and Organization of this General Discussion Chapter

Although the chapters that comprise this thesis all relate to the broad topic of systematic reviews and CAM, the component chapters have been divided into two 'Parts' for this thesis. The reason that these two sets of studies, or 'Parts', have been examined separately in this thesis is because these two Parts addressed largely different objectives. The main objectives addressed in Part 1 (i.e., Chapters 2-4[1-3]) were the development and analyses of trial and systematic review evidence bases on CAM. Because the three studies that comprise Part 1 are discrete, and because each of these three studies has different implications for research and practice, these three studies are each discussed separately below, each under its own section heading, in Part 1 of this General Discussion Chapter. The main objectives addressed in Part 2 (i.e., Chapters 5-9[4-8]) were the preparation of systematic reviews to evaluate the effects of acupuncture for three specific conditions, and the conduct of research with a view to improve the methods of future acupuncture RCTs and systematic reviews. In contrast to Part 1, in which the three section headings correspond to the three *individual component studies*, in Part 2, the section headings are organized according to a summary and discussion of the *common methodological issues* encountered in preparing all three included systematic reviews of acupuncture.

Part 1. Development and Analysis of CAM Systematic Review-related Databases: Reflection on the Main Findings and Primary Methodological Issues Encountered

Although there is widespread use of CAM therapies by the Public, one barrier to the integration of CAM therapies into the healthcare system has been the lack of a systematic scheme for compiling and organizing the highest quality CAM research studies, and classifying the bottom-line findings of these studies. In Chapter 2, an 'operational' definition of CAM and a CAM topic classification scheme was developed that was used to identify, classify, and organize all CAM-related Cochrane reviews included in Issue 4, 2009 of *The Cochrane Library*, so that these reviews could be disseminated through a Cochrane CAM reviews 'topics list'. This CAM reviews 'topics list' included 396 CAM-related Cochrane reviews as of Issue 4, 2009 of *The Cochrane Library* (October 2009), but this number has since increased to 596 CAM-related Cochrane reviews as of Issue 12, 2012 of *The Cochrane Library* (published in December 2012, with the change from quarterly to monthly publication). While a substantial proportion of the CAM-related Cochrane reviews on the topics list focus on acupuncture and other traditional Chinese medicine (TCM) topics (i.e., 111/596 (18.6%) of all CAM-related Cochrane reviews, as of Issue 12, 2012), the topics list also includes Cochrane reviews of all other types of CAM interventions (e.g., probiotics, massage, chiropractic etc.) Though initially developed for the CAM reviews topics list, the operational definition of CAM and the CAM topic classification scheme were also used to identify, organize, and compile the 44,840 records of CAM-related controlled trials included in the CAM trials database (Chapter 3). As described in Chapter 3, the improved accessibility and availability of these typically difficult to locate, and often low quality, trials included in the CAM trials database presents both opportunities and challenges for future systematic review authors in CAM. Finally, in Chapter 4, an effort is presented to classify the bottom-line findings of all TCM related Cochrane reviews, in terms of whether or not the reviews' conclusions suggested a benefit of the TCM intervention. As discussed

below, in the context of Chapter 4, it was found that classifying the bottom-line findings of systematic reviews is a complex endeavor, which requires additional research. However, the methods used and the challenges encountered in categorizing these TCM-related reviews' overall findings may serve to inform future research efforts to classify the overall findings of systematic reviews.

1.1. Developing an Operational Definition of CAM (Chapter 2)

Why is developing an operational definition of CAM a topic worthy of researching? After all, some leading authorities have argued that there is no CAM, but only health care interventions, and that defining interventions as CAM or conventional is an artificial distinction.[9] Others have questioned whether it makes sense to define an overall umbrella category of CAM that encompasses interventions that are homogeneous in multiple ways.[10,11] For example, homeopathic remedies and probiotics, two interventions that might be grouped under this umbrella category of 'CAM', have different levels of demonstrated efficacy, different levels of plausibility for their putative mechanisms of effect, and different levels of acceptance. Despite this heterogeneity of CAM interventions, developing an operational definition of CAM was still believed to be a worthwhile project because CAM is a widely used term among the medical research community. Indeed there is an entire US National Institutes of Health Center dedicated to studying CAM. Therefore, to discuss and use the term CAM, it needs to be defined.

On a pragmatic level, a transparent and reproducible operational definition of CAM was needed for developing the topics list of all CAM-related Cochrane reviews, grouped by CAM intervention topic areas, as described in Chapter 2. This topics list has been valuable in allowing for the tracking of CAM-related Cochrane reviews, and for facilitating the accessibility of these reviews to journal editors, governmental organizations, physician groups, patients, and researchers. By preparing and keeping this topics list current, requests for information about Cochrane CAM-related reviews can be quickly addressed, providing those who request the information with a current list of CAM-related Cochrane reviews, selected using explicit, transparent, and reproducible criteria. Links to the Cochrane CAM Field topics list are now included on the web pages of the US National Institutes of Health, National Center for Complementary and Alternative Medicine; the International Society for Complementary Medicine Research; and the US Consortium of Academic Health Centers for Integrative Medicine. To assist other Cochrane Fields in developing topics lists related to their own fields of research, workshops have been provided at three previous Cochrane Colloquia in which the challenges encountered in developing the CAM Field topics list were discussed.

The operational definition of CAM developed for Chapter 2 has also been used for defining the eligibility criteria for the two other thesis chapters related to CAM trial and review database development (i.e., Chapters 3 and 4), as well as for other studies not included in this thesis. For example, our CAM operational definition was used to determine whether (or not) a given trial should be considered CAM-related, and thus eligible for inclusion (or not), in a recent study of the completeness of safety reporting in CAM trial publications.[12] As another example, our CAM operational definition was used to determine the eligibility criteria (i.e., CAM or not CAM) in a recent systematic review of cost-effectiveness studies of CAM interventions.[13] Finally, our operational definition of CAM may facilitate CAM research in areas unrelated to either database development or systematic reviews of CAM studies. Namely, as noted in the General Introduction Chapter, comparisons of survey research on the prevalence of CAM use, over time and across countries, has to date been hampered by the fact that different definitions of CAM have been used across different surveys, perhaps because no comprehensive, operational definition of what should be considered CAM had heretofore been available. Our operational definition of CAM may provide a standardized definition of CAM to use for assessing CAM use in future surveys.

1.2. Developing and Analyzing a Database of CAM-related Controlled Trials (Chapter 3)
Chapter 3 describes the methods for developing a database of CAM-related controlled trials, as well as a bibliographic analysis that categorized the characteristics of the 44,840 trial records included in the database, according to CAM intervention types evaluated, language of publication, MEDLINE availability of the records, etc. This bibliometric analysis found that the CAM intervention type associated with the largest numbers of trials in the database was non-vitamin, non-mineral natural products (n=15,140), followed by Chinese herbal medicine (n=12,118). Many of the trials included in the database had previously been inaccessible to systematic review authors. For example, of the 44,840 CAM trial records in the database, 36% are not included on MEDLINE, and 29% are not reported in English. Of the Chinese herbal medicine trial records, 71% are not included on MEDLINE. Therefore, the CAM trials database may be a particularly useful resource for identifying 'difficult-to-locate' trials (i.e., those that are not English language and/or are not included in MEDLINE) for systematic reviews of traditional medicine therapies. However, this strength of the register may be associated with potential weaknesses in terms of the quality of included trials. For example, the largest subset of non-MEDLINE trials in the database is Chinese-language trials of traditional medicine. Many of these trials have not used adequate methods of randomization,[14] have high risks of bias,[15] and almost always report positive findings.[16] The Discussion section of Chapter 3 suggests some potential ways that systematic review authors may proceed when confronted with these difficult to locate trials, including how review authors may address potential reporting and publication biases associated with these trials.

1.3. Overview of TCM-related Cochrane Reviews (Chapter 4)
Chapter 4 describes a review overview of all Cochrane reviews on the topic of traditional Chinese medicine (TCM). One of the objectives in preparing this review overview was to determine which, if any, TCM therapies have suggestive evidence of a benefit according to Cochrane reviews. This is important because identifying Cochrane reviews of TCM therapies with suggestive, but not yet conclusive, evidence of efficacy might aid in prioritizing new RCTs, and updates of Cochrane reviews, on these specific topics. An initial challenge encountered in preparing this review overview was deciding which Cochrane reviews should be considered TCM-related, and thus eligible for inclusion in this review overview. This challenge was relatively easily addressed, by referring to the operational definition of CAM that had been developed (as described in Chapter 2), which includes categories related to TCM.

The next challenge that was encountered, which was less easily addressed, was evaluating and classifying the findings of the included TCM-related Cochrane reviews. Because no guidelines or recommendations existed for systematically, reproducibly, and transparently evaluating the bottom line findings of a large number of systematic reviews, it was necessary to devise a methodology for doing so. For evaluating the bottom line findings of the strength of the evidence for the 70 TCM-related Cochrane reviews included in this review overview, the decision was made to rely largely on the 'Authors' Conclusions' section of the systematic reviews' abstracts. That is, transparent and explicit criteria were developed to classify the Authors' Conclusions into one of the following two categories: 1) suggestion of a benefit of the TCM intervention, which was qualified by a caveat about the poor quality and quantity of included studies, or 2) no conclusions could be drawn about the TCM intervention from the currently available data. A category for evidence for absence of effect was not included because no Authors' Conclusions concluded that there was firm evidence of inefficacy of the TCM intervention. This is not surprising because these systematic reviews of TCM interventions generally did not include a sufficient number of large, high quality, homogeneous RCTs to conclusively determine an *absence* of effect. Similarly, the unqualified Authors' Conclusions of a clear *presence* of an effect could also not be drawn for any of the reviews because of the typically

low quality and heterogeneous trials included. Finally, a category for 'harm' was also not included because the Authors' Conclusions were examined before deciding on the two categories, and none concluded that the intervention was harmful. However, this does not mean all of these TCM interventions are safe. Indeed, adverse events are often not well-reported in CAM RCTs,[12] and a systematic review of RCTs is not the optimal design for identifying rare but serious harms.[17] On the other hand, a systematic review of large cohort studies could theoretically be a very strong design to study harms.[18]

Our assessment of the Authors' Conclusions was also supplemented with a listing of the comparisons and outcomes showing statistically significant meta-analyses results. These statistically significant results were listed only for those reviews for which the Authors' Conclusions suggested, but did not firmly conclude, that there was evidence of efficacy (i.e., "category 1" reviews). The reason for this decision was that if the Authors' Conclusions indicated that no conclusions could be drawn based on the evidence, then it was assumed that any statistically significant meta-analysis results were likely generated from methodologically unsound trials. Indeed, a disadvantage of relying *exclusively* on the statistical significance of the meta-analyses results (which is often the approach used for evaluating a large number of systematic reviews for evidence-based clinical practice guidelines[19,20]) is that it does not address the biases and flaws of the included trials, whereas the Authors' Conclusions should take into consideration not just the numerical effect estimates, but also the quality of the evidence.

As noted above, in preparing this review overview, it was found that there were no guidelines or recommendations for Cochrane review authors for writing up Authors' Conclusions statements. As a result, there has been no consistent approach among TCM review authors for translating the strength of the reviews' findings into standardized Conclusions. Therefore, different review authors might draft different Conclusions based on a review of the same trial data, thereby presenting a major obstacle for preparing review overviews informed by the systematic reviews' Conclusions. Bias can also arise in the step from Results to Conclusions, [21] as illustrated by a review of reviews that compared Cochrane reviews with industry-supported meta-analyses on the same drug.[22] This review of reviews found that while the Cochrane reviews and industry-supported meta-analyses had similar estimated meta-analytic treatment effects, the industry-supported meta-analyses had more favorable conclusions about the drug than the corresponding Cochrane reviews, largely because the Cochrane reviews' Conclusions had more often considered the potential for bias in the reviews. Based on their findings, the authors of this review of reviews noted in their Discussion section that "This study suggests that the main problem in industry-supported reviews lies in how conclusions are formulated." The conclusion of some Cochrane reviews can also be too optimistic, although the problem is much less pronounced than for industry-supported reviews.[22] For example, an analysis of 53 Cochrane reviews from 1998 found that the evidence did not fully support the Conclusions in 17% of Cochrane reviews, and in all cases the problematic Conclusions were found to have been too optimistic.[23]

The incorporation into Cochrane reviews of Summary of Findings tables, with their use of the GRADE approach to grade the quality of the evidence for each outcome in a review in a transparent and systematic way, has been an important development for using a systematic and standardized approach to communicate the findings of Cochrane reviews in a tabular format.[24] One future area of research might be to investigate whether the GRADE findings could be incorporated or translated into more complete and detailed systematic review Authors' Conclusions statements. However, if this were to be considered, it would be important to balance the need to use a systematic, standardized, and unbiased approach in drafting Authors' Conclusions, with the simultaneous need to allow authors the freedom and flexibility to briefly describe these Conclusions with a

wording that most precisely and accurately encapsulates the evidence. It might be possible to achieve a middle-ground which addresses both of these concerns.

In preparing this review overview, it was also found that the Authors' Conclusions were often very brief. More detailed and precise Conclusions may better serve many readers, who rely on only these abstract Conclusions to learn what the review has found. Indeed, brief Conclusions about whether or not a TCM treatment "works" may be unrealistic. The reason for this is because systematic reviews of TCM have typically included trials with different variations of the treatment, and the treatments have been compared to a wide variety of controls. The results of these different treatment/control comparisons may differ. Longer Conclusions would provide review authors the space to provide a more nuanced, detailed, and complete interpretation of these complex results. Indeed, the Cochrane Collaboration has recently conducted an audit of Cochrane review abstracts, and the first recommendation resulting from this audit was that Cochrane review abstracts should be increased from 400 words to a maximum of 1000 words (although abstracts no longer than 700 words are encouraged).[25] The audit noted that a consequence of the current brief abstracts "is the loss of information that would otherwise provide a more complete picture about the review question and findings." The Cochrane audit also concluded that brief abstracts can be especially problematic for multiple comparison reviews, which is the case for many TCM-related reviews.

While more detailed and standardized Authors' Conclusions might facilitate the preparation of review overviews, another approach may be to instead use the Plain Language Summaries (PLS), generated from the Summary of Findings (SoF) tables, rather than the Authors' Conclusions, as the basis for preparing future review overviews. Because these SoF/PLS statements are comparison/outcome specific, and because they are generated using transparent and systematic methods, the SoF/PLS statements may be potentially less susceptible to bias and may also represent more "concrete" conclusions than the Authors' Conclusions statements. These SoF-based PLS use 'plain language' to translate the findings of the quality and strength of the evidence for each comparison/outcome, from the SoF table, into easy to understand, standardized statements. That is, in SoF-based PLS, different wording corresponds to different "grades of evidence": the words *will* (high-quality evidence), *probably* (moderate-quality evidence), and *may* (low-quality evidence) describe the certainty about the effect of the intervention for each comparison/outcome in a review.[26] While most Cochrane reviews did not include a SoF/PLS at the time that Chapter 4 was prepared, these SoF/PLS are increasingly being included in Cochrane reviews. Should this review overview be updated, an alternative approach for evaluating the Cochrane reviews' bottom-line conclusions that may be preferable to the abstract conclusions statements would be using the SoF/PLS.

The review overviews project initiated by the Cochrane Collaboration may also in the future address the need for a transparent and scientific method to summarize systematic reviews of a class of interventions. However, as of November 2012, only six such resource-intensive review overviews have been prepared, and they have each used different approaches.[27-32] Three out of six of these review overviews restricted eligibility to Cochrane reviews;[27,30,32] the six review overviews generally included only a small number of systematic reviews (i.e., median six, as opposed to the 70 Cochrane reviews that were summarized in Chapter 4); and the review overviews used a variety of different methods to summarize the included systematic reviews' findings. For example, two of these review overviews[29,31] summarized the data from the included systematic reviews using a narrative or "vote counting" approach. The authors of these two review overviews[29,31] acknowledged that while vote counting has limitations, they used this method because there are few other robust alternative methods to summarize such diverse results across systematic reviews. Two other review overviews used network meta-analyses to provide quantitative estimates of direct and indirect comparisons between different types of treatments for a specific condition.[27,32] Thus, while the Cochrane

Collaboration's review overviews project is extremely valuable, it may not yet, on its own, entirely address the need for a way to summarize a large number of Cochrane and non-Cochrane systematic reviews, using a standardized methodology.

Part 2. Systematic Reviews of Acupuncture and Related Methodological Research: Reflection on the Main Findings and Primary Methodological Issues Encountered

The included systematic reviews of acupuncture for osteoarthritis and low back pain (Chapters 7 and 8) found that acupuncture has only small reported benefits relative to a sham acupuncture control, and that even these small reported benefits may be due at least partially to placebo effects from incomplete blinding. However, both true acupuncture *and sham acupuncture* typically showed clinically relevant benefits compared to a waiting list control and some standard medical care controls. These findings suggest that either acupuncture has a particularly potent placebo effect for the subjective outcomes evaluated (e.g., pain), or that some of the sham acupuncture interventions that have been used as controls in RCTs may have affected the outcome evaluated, through the same putative mechanism of true acupuncture. The systematic review of acupuncture as an adjuvant to IVF (Chapter 9), for which the primary outcome was entirely objective (e.g., clinical pregnancy) and therefore much less influenced by placebo effects, found that there were no pooled benefits of adjuvant acupuncture across all included RCTs. However, adjuvant acupuncture appeared to increase pregnancy success rates in the subgroup of RCTs with lower baseline pregnancy rates, where baseline rates were measured as the observed rate of pregnancy in the control group of each trial. In preparing these three systematic reviews, a main objective, as with any systematic reviews, was to determine the effectiveness of acupuncture for the given conditions for which it was evaluated in order to inform clinical practice guidelines and policy decisions. Another important objective was to identify limitations of the RCTs included in these reviews, in order to provide recommendations for the design of future RCTs. To adequately address these objectives, it was necessary to first learn as much as possible about the RCTs reviewed, often by contacting the original RCT authors. It was next necessary to conduct a thorough and rigorous analysis of the information obtained, in order to best estimate the truth of the effects of acupuncture for the given conditions. Each of the major methodological issues that were encountered in preparing these systematic reviews (e.g., contacting trial authors, conducting subgroup analyses, determining clinical relevance) is discussed separately, below. Part 2 concludes with a brief, general overview of implications for future research of systematic reviews of acupuncture.

2.1. Contacting Trial Authors to Obtain Missing Data in Systematic Reviews

The ability of systematic review authors to draw robust and specific conclusions that can best guide clinical practice, as well as best guide the design of future RCTs, depends on the review authors having access to adequate information on the design, conduct, and findings of each of the RCTs included in the review. However, the relevant information needed to make these assessments is often missing from RCT publications. This missing information, on risk of bias dimensions, for example, presents a dilemma for systematic review authors. Namely, review authors must decide whether to contact the RCT authors to obtain the missing data related to risk of bias, or whether, instead, to assume that the methodological details related to risk of bias not described in the publication were most likely not part of the RCT's procedure. The issue of contacting RCT authors for additional information is a theme that reoccurs through all three of the systematic reviews in this thesis (i.e., Chapters 7-9). In addition, a study was conducted specifically to address the issue of the value of contacting RCT authors for additional data (i.e., Chapter 6). Contacting RCT authors for further information about their RCTs is one of the most labor-intensive and time-consuming

components of preparing a systematic review, but it may also strengthen the validity of review findings.

In making efforts to directly contact RCT authors to request relevant information not reported in their publications, two major issues were identified. The first was deciding which, and how much, unreported information a review author should attempt to collect. The second was deciding the best method of asking the relevant questions to collect this information. First, in terms of which, and how much, information to collect from RCT authors, experience with the reviews included in this thesis has led to the conclusion that a judicious selection of the most important and relevant information should be prioritized for collection. That is, in contacting RCT authors, review authors should request information that has the potential to impact the overall results and conclusions of the systematic reviews. This includes information necessary for the calculation of robust effect estimates for the meta-analyses, and information necessary for a thorough risk of bias assessment of the included RCTs. The issue of obtaining relevant data for calculating robust effect estimates was most relevant for the acupuncture versus sham comparisons. For the acupuncture versus sham comparison in the osteoarthritis review (Chapter 8), for example, the standardized mean differences (SMDs) were small and therefore potentially sensitive to the use of different methods of calculating these SMDs. Therefore, the SMDs for this comparison were calculated using both changes from baseline (as the primary analysis) and post-treatment values (as a sensitivity analysis). However, because only one or the other sets of values were typically reported in the RCT publications, the unreported sets of values needed to be collected from RCT authors. Collecting this data increased the time and effort involved in preparing the osteoarthritis review, but also increased the rigor of the final meta-analyses because few assumptions or imputations regarding the RCTs' outcome data needed to be made. In terms of risk of bias domains and clinical characteristics, there was a focus on requesting information from RCT authors only on the specific risk of bias domains and clinical characteristics that were to be used for pre-specified subgroup analyses. However, in addition to asking specific questions of RCT authors to obtain this most critical information, the RCT authors were also provided with the risk of bias assessments table (together with its scoring criteria) and the characteristics of included studies table for their RCTs. These tables were provided in the event that the RCT authors were also available to go through these tables, and provide any further supplementary information not reported in their RCT publications.

As mentioned above, the second important consideration in obtaining unpublished information from RCT authors is determining the best methods of framing the questions for requesting the most relevant information. The wording and format used to ask the specific questions of the RCT authors was challenging because it required the need to collect specific and detailed information, while at the same time, to avoid question formats that might lead to a response bias. In the initial e-mails to RCT authors, general open-ended questions were asked (e.g., about randomization method), rather than asking the investigators to select from a list of response choices, to avoid a possible positive response bias. Some of the subsequent follow-up questions needed to be more specific to obtain more specific and detailed information. While it was assumed that open ended questions would result in a more truthful and valid response than response choice questions, a study to corroborate this assumption may be useful.

Finally, in contacting the RCT authors to learn the precise methods of randomization, it is of concern that no matter how this question is asked, the question may not include sufficient safeguards to insure a completely accurate and truthful response. Therefore, a more specific question, and one which would require the RCT authors to explain how a result reported in their publication came about, would be helpful. One such specific question that was identified as useful was asking the RCT authors to explain how they achieved identically sized treatment groups, when this was the case. The response to this question could be revealing because if an RCT author informs

a review author, in response to an initial contact, that a simple randomization scheme was used (e.g., referring to a random number table), but then was unable to later explain how this procedure achieved identical numbers in each group, it raises the possibility that the randomization might not have been adequately generated or concealed, or that some patients were non-randomly assigned. Indeed, the (wrong) notion that RCTs should have equal numbers in each treatment group has been shown to commonly lead RCT investigators to force equality by unscientific means.[33] In fact, previous methodological reviews of this issue have found that over one-half of RCTs using simple, unrestricted randomization schemes report identical numbers in each group.[33,34] This additional question requesting an explanation of how identical group sizes were achieved was found to be especially useful for Chinese trials, where there is already great uncertainty about whether trials reported to be randomized in their publication text are authentic randomized trials.[14] A description of how identical group sizes were achieved in trials using simple randomization might be considered for inclusion in a future update of the CONSORT statement.

If the CONSORT statement[35] and its corollary guideline for reporting of treatment-related components in acupuncture RCTs, the STRICTA statement,[36] were fully adhered to by acupuncture RCT authors, the relevant information needed by acupuncture systematic review authors would largely be included in the RCT publications. This would reduce the need of systematic review authors to contact RCT authors, often many years after an RCT is published, at which time the relevant information may be difficult or impossible to retrieve. Indeed, if an RCT's corresponding author cannot answer a systematic review author's questions (e.g., about how identical group sizes were achieved), it does not necessarily mean that there is bias, because a statistician, who may no longer be accessible, may have been responsible for this part of the RCT's design and implementation. Although the CONSORT statement has been a tremendous advance in improving the reporting in RCT publications, to further maximize systematic review authors' access to complete and accurate information about the methods and results of the RCTs included in systematic reviews, two regulations would be necessary: 1) requiring the separate (prior) publication of the full RCT protocol, with all relevant information on the methods used; and 2) requiring the publication of the data set used for the analyses that lead to the tables in the published RCT article. Once such regulations are in place and consequently all relevant information about RCTs are available in the public domain, extensive contacts with authors of RCTs will no longer be necessary, future systematic reviews will be less time-consuming to prepare, and systematic reviews' findings may also be more valid.

Six studies, including ours (Chapter 6), have assessed whether the methodological quality of RCTs as reported in RCT publications, reflects their actual methodological quality. The largest and most recent of these studies, by Mhaskar and colleagues[37] included 429 phase III RCTs from US National Cancer Institute Cooperative Groups. This 2012 study compared the reported quality, based on the RCT publication, with the actual quality, based on the RCT protocol. This study found that poor quality of RCT reporting, as based on the publications, does not reflect the actual high methodological quality of RCT conduct, as based on the protocols. This Mhaskar and colleagues publication also included a summary of previous studies, and concluded that their own study's findings were "in line with previous research on the topic". Thus, the studies conducted to date suggest that it is indeed important to contact RCT authors to obtain important methodological details that are unreported or inadequately reported in RCT publications, and that it cannot be assumed that methodological details not described in the publication were not part of the RCT's procedure.

As discussed above, contacting RCT authors is necessary to obtain additional, unpublished information about the RCTs included in a systematic review, primarily so that review authors will have adequate data to conduct subgroup and sensitivity analyses. In the next section, the issue of

subgroup analyses is discussed, and for acupuncture systematic reviews, the most important of these, which is the subgroup analysis on adequacy of patient blinding.

2.2. Subgroup Analyses

Systematic reviews involve an attempt to synthesize and summarize the findings of studies that often have heterogeneous methods. For example, different RCTs included in a review may have different risks of bias (e.g., differential blinding of patients, differential completeness of outcome data) and these bias-related dimensions need to be evaluated in synthesizing and summarizing the results of the RCTs. Also, because acupuncture (and CAM more generally), is more difficult to standardize than pharmaceuticals, there is often great heterogeneity in the treatments administered across acupuncture RCTs, as well as heterogeneity in the adequacy of these treatments. Combining the results of all RCTs without a careful consideration of these differences in risks of bias and treatment adequacy variables would be an erroneous over-simplification. Therefore, subgroup analyses are important for comparing the effects of treatments across subsets of trials. However, before the issue of subgroup analyses is discussed more generally, a discussion is first provided of the most important subgroup analysis, related to the most important quality criterion for acupuncture reviews, which is the adequacy of RCT participant blinding. Indeed, in the acupuncture for osteoarthritis review (Chapter 8), the effects of acupuncture versus sham differed depending on whether or not the sham control adequately blinded RCT participants.

2.2.1. Evaluating adequacy of sham acupuncture controls for use in subgroup analyses.

In the systematic reviews of acupuncture for low back pain and osteoarthritis (Chapters 7 and 8), the outcomes evaluated were entirely patient-reported, e.g., pain and function. Because such outcomes are self-reported and subjective, outcomes assessment bias is a major threat to validity in these two reviews. That is, the participants in these RCTs who knew they were receiving acupuncture might have expected acupuncture to benefit them more than usual care therapies. As a result, the acupuncture group participants may have reported feeling less back or knee pain on the outcomes assessment questionnaires, regardless of whether the acupuncture truly worked. To control for such expectation effects, some of the RCTs in these reviews used sham acupuncture controls to blind RCT participants as to whether or not they were receiving true acupuncture or a sham. However, if the sham controls are not sufficiently believable to RCT participants to adequately ensure blinding to treatment assignment, then uncontrolled expectation effects may threaten the validity of the results of sham-controlled RCTs that show a benefit of true acupuncture over sham. Therefore, in the low back pain and osteoarthritis reviews, whether or not the sham controls were believable to participants, and the effect this may have had on the RCTs' results, was carefully examined.

However, developing a believable, yet physiologically inert, sham acupuncture control is not as straightforward as developing a placebo pill, i.e., the typical control used in drug RCTs. The fundamental challenge for acupuncture RCT investigators has been developing a sham acupuncture control that is sufficiently believable to participants so as to be indistinguishable from true acupuncture, and yet, at the same time, not so similar to true acupuncture that the sham has a specific therapeutic effect of its own, through the same putative mechanism of true acupuncture.[4,38] In other words, because we do not know the precise mechanism through which acupuncture 'works' to exert its putative effect, it is very difficult to know what to mimic and what to avoid, in designing the sham acupuncture control intervention. One way to address this difficulty, moving forward, is for basic science researchers to systematically evaluate the specific effects of each of the components of true acupuncture treatment (e.g., insertion depth, needle location, needle stimulation) so that the sham control can avoid mimicking acupuncture-specific effects due to each

of these components.[39] In addition, if basic researchers were to conclusively show that acupuncture points are unique anatomical structures, which has not been conclusively shown to date,[39,40] there would be clearer criteria for selecting 'non-acupuncture points' to use in future sham acupuncture interventions.

Acupuncture systematic review authors are faced with several corresponding challenges related to sham-controlled RCTs: 1) developing criteria for assessing whether the sham control interventions used in the RCTs reviewed were believable to RCT participants, yet also lacking acupuncture-specific effects; 2) incorporating these assessments of the sham controls into the review's analyses; and 3) taking the analyses of the sham control assessments into consideration in drawing overall conclusions. In the three systematic reviews included in this thesis (Chapters 7, 8 and 9), progressively more sophisticated methods were used for evaluating both the risk of acupuncture-specific effects of the sham controls used, and the likely credibility of these sham controls among RCT participants, because the importance of each of these two variables was increasingly recognized.

2.2.1.1. Assessing the risk of the sham interventions having acupuncture-specific effects.

For the first review, on acupuncture for low back pain (Chapter 7), no formal assessments were made of the risk of the sham interventions affecting the outcome (e.g., pain) through the same putative pathways of true acupuncture (i.e., stimulation of the true acupuncture points; needle penetration of the skin). For the second review, on acupuncture for osteoarthritis (Chapter 8), two experienced acupuncture practitioners/researchers, who were blinded to the results of the RCT and the publication, each classified the sham interventions according to whether or not these sham interventions would have a risk of having an acupuncture-specific effect on the outcome. The acupuncturists made these assessments independently, with a consensus achieved by discussion. Because this method relied on the assessors' subjective judgments, it may have resulted in considerable inter-observer variation or inconsistency of assessments across trials. To improve the reproducibility of these assessments, for the acupuncture as adjuvant to IVF review (Chapter 9), the subjective method of assessment used in the osteoarthritis review was used, as well as a more objective method of classification. For the more objective evaluation method, the sham controls were dichotomously classified on whether or not they had a high likelihood of acupuncture-specific effects,[38] based on whether or not the needles were inserted under the skin, and where the needles were placed relative to the acupuncture points.[41]

2.2.1.2. Assessing the credibility of the sham acupuncture controls.

In the three reviews, increasingly sophisticated methods were also used for assessing whether the sham acupuncture control intervention was sufficiently believable to RCT participants as an authentic treatment, such that the sham would fully allow for the control of placebo effects. For assessing whether the sham control was believable to the RCT participants, in the first review on low back pain (Chapter 7), an additional quality assessment variable was added on whether or not the success of participant blinding was tested after treatment. To improve upon this simple method used for assessing sham control credibility, for the acupuncture for osteoarthritis review (Chapter 8), the sham control credibility was assessed using two different methods, which allowed for greater flexibility and precision of these assessments. That is, for the osteoarthritis review, the sham controls were considered to be credible if *either* one of two different credibility criteria were met. First, an RCT was considered to have used a credible sham if the RCT used needle acupuncture as the sham *and* the RCT's investigators informed participants that two different types of acupuncture were being compared (i.e., the investigators did not inform RCT participants that they might receive a sham treatment). This first credibility criterion was decided upon because a sham acupuncture

intervention that involves needle insertion is likely to be believable to participants as true acupuncture, particularly if it is not disclosed to the RCT participants that they might be receiving a sham treatment. However, this non-disclosure practice may not be permitted by some internal review boards, and some bioethicists[42] have argued that this practice violates the ethical requirements of informed consent. Second, an RCT was considered to have used a credible sham if the credibility of the sham was evaluated, and found to be indistinguishable from true acupuncture. This evaluation procedure involves the RCT investigators asking the RCT participants to guess whether they received an authentic treatment or a sham.[43] For RCTs in which the sham-like comparison intervention is not described as a sham during the consent procedure, the credibility test has instead involved asking RCT participants to guess whether they thought "they had received acupuncture following the principles of Chinese medicine or the other type of acupuncture."[44,45]

Although this sham credibility evaluation procedure was used in the reviews, the validity of this procedure of testing the credibility of the sham is unclear and deserves further study because correctly guessing treatment assignment could be highly correlated with a treatment's effectiveness, or lack thereof.[46-48] That is, if only true acupuncture and not sham acupuncture had analgesic effects, for example, then one might expect that RCT participants assigned to true acupuncture would be more likely to guess that they were receiving an authentic treatment than would participants assigned to the sham acupuncture, because only the true acupuncture participants would be experiencing the analgesic effects of the treatment. Or as Altman and colleagues describe it, "end of trial tests of blindness might be tests of hunches for adverse effects or efficacy".[47] Indeed, both Altman and colleagues[47] as well as Sackett[46,49] have argued that RCT investigators should vigorously test for blinding success before RCTs start, but that tests of blinding success at the end of RCTs can be difficult to interpret, and therefore cannot be recommended in all circumstances. Despite these concerns that had been raised about testing the credibility of the sham in this way, the reason that tests for success of blinding after the RCT had begun were used in the included reviews is because acupuncture is a particularly difficult intervention for which to develop a believable sham control, and it is possible that the RCT participants would continue to try to figure out whether they were receiving the true or sham acupuncture intervention while the RCT was underway.

Since the acupuncture for osteoarthritis Cochrane review (Chapter 8) was published in 2010, further guidance stressing even greater caution in evaluating the results of RCT participants' guesses to treatment assignment have been published, for example in the 2010 CONSORT statement[35] and its corollary publication,[50] in which the mention of how the success of blinding might have been evaluated has been removed.[35,50] In an effort to more appropriately address these concerns about the validity of RCT participant guesses of treatment received, and also to use more precision in how this variable was defined, for the next update of the acupuncture for osteoarthritis Cochrane review, the time point for assessing the results of patients' guesses as to treatment received has been defined as the point closest to four weeks after treatment has begun. This four week time point was decided on because it was judged that any differences in tests of RCT participants' guesses after four weeks could be influenced by differences in treatment efficacy between acupuncture and sham. On the other hand, differences before four weeks may be more likely due to participants figuring out whether they received the true or sham intervention based on their observations of the intervention protocol used. The four week time point was also selected because four weeks is typically the earliest time point that participants' guesses to treatment assignment is assessed in trials. However, if true acupuncture has a better efficacy than sham acupuncture earlier than four weeks after treatment initiation, then even this four week time point for assessing participants' guesses to treatment assignment could potentially be influenced by the (lack of) specific effects of treatment.

As described above, assessing the adequacy of RCT participant blinding was probably the most important methodological consideration in the low back pain and osteoarthritis reviews (Chapters 7 and 8). Because both of these reviews used subjective, participant-reported outcomes (i.e., pain and function), participant blinding was important to obtain unbiased outcomes assessments. That is, ratings of pain and similar subjective states can be strongly influenced by respondents' prejudgments, preferences, and expectations about treatment benefits. Therefore, controlling for these placebo effects due to patient expectations, by using a sham acupuncture control group, is critically important for preventing biased participant reporting of outcomes in RCTs of acupuncture for pain-related conditions. However, in the third systematic review included in this thesis, which evaluated acupuncture as an adjuvant to IVF (Chapter 9), the outcomes evaluated are entirely objective (i.e., clinical pregnancy, ongoing pregnancy, and live birth). In such RCTs with entirely objective outcomes, biased participant reporting of outcomes is not an issue. Therefore blinding RCT participants to treatment assignment is unnecessary for this reason.[38,51-53] However, in addition to this first reason of preventing biased participant outcome reporting, a second, completely different reason for using a sham control to blind participants to treatment assignment is to distinguish the specific effect of the point-specific placement of the acupuncture needles from any non-specific psychobiological placebo effects due to patient expectations, the treatment setting, patient-practitioner interaction etc. This second reason for RCT participant blinding was relevant not just for the osteoarthritis and low back pain reviews, but also for the IVF review. At first glance, it may seem unlikely that an IVF RCT participant's knowledge of whether she was receiving adjuvant acupuncture would affect her ability to become pregnant from IVF; however, it is indeed possible that adjuvant acupuncture may increase IVF success rates strictly through a non-needling-related psychobiological placebo effect mechanism. This may occur, for example, if women who believe they are receiving true acupuncture (whether or not they truly are) have reduced stress, which thereby subtly influences their pregnancy-related hormone production. Therefore, using a sham control in IVF RCTs may help elucidate the mechanism responsible for adjuvant acupuncture's effects on IVF success rates. However, even if adjuvant acupuncture were to increase IVF success rates strictly through a psychobiological placebo effect stress reduction mechanism, this effect would still be integral to the working mechanism by which adjuvant acupuncture increases IVF success rates. Therefore, for addressing the more clinically relevant question of the total effects of acupuncture (i.e., specific needling effects plus any non-needle-related psychobiological placebo effects) in contributing to any increases in IVF success rates, a sham control seems unnecessary.[38]

In the sections below, it is described how subgroup analyses were used in the included reviews to assess the impact on the RCTs' outcomes of the risk of the sham controls having acupuncture specific effects; the credibility of the sham controls; as well as the impact of other risk of bias- and treatment adequacy-related variables.

2.2.2. Subgroup analyses and risk of bias.

Because inadequate patient blinding, as well as several other RCT risk of bias components, may be associated with exaggerated treatment effects,[53] it is a standard and universally accepted practice that a systematic review author must evaluate risk of bias components of the RCTs included in a systematic review. These evaluations can then be used by review authors to conduct subgroup analyses to test whether specific risks of bias are associated with more or less positive RCT outcomes. The impact of each risk of bias component on outcomes can be assessed individually, or by using a summed composite score. In the three reviews included in this thesis, the individual component method was used, primarily because it is the approach advocated in the Cochrane Handbook.[54] By using the individual component method, it was possible to tease out which individual risk of bias components were associated with treatment success. For example, in the

acupuncture for osteoarthritis review (Chapter 8), it was determined that the individual risk of bias variable of adequacy of RCT participant blinding predicted treatment success. However, if the summed composite score approach had been used instead, this adequacy of participant blinding variable would have been mixed in with the other disparate risk of bias variables into a single uninformative, opaque composite score. And using such a composite score for analysis on quality would not have addressed or detected the impact on outcomes of the individual risk of bias component of adequacy of patient blinding. However, a disadvantage of the individual component method is that most systematic reviews include only a small number of RCTs,[55] and consequently there is inadequate power to conduct multiple subgroup tests to analyze for the effects of a number of different individual risk of bias components on outcomes.[56] By using a summed composite score to analyze for the effect of overall risk of bias on outcomes, review authors do not need to conduct separate subgroup tests on multiple risk of bias components. In addition, a recent study[57] suggests that a summed score may indeed adequately distinguish between RCTs with an overall high versus low risk of bias.

2.2.3. Subgroup analyses and treatment adequacy.

Study quality instruments (e.g., the Cochrane risk of bias tool, the Cochrane Back Review Group checklist) evaluate risk of bias elements (e.g., adequacy of randomization or blinding). Trials with a high risk of bias can have results that are biased, usually, although not always, on the side of a type I error (a false-positive finding). However, in addition to evaluating risk of bias elements, it is also important for systematic review authors to evaluate elements related to the adequacy of the acupuncture treatment procedures in the included trials. Such an evaluation is important because most acupuncture interventions (and CAM interventions in general) lack the "dose-finding" Phase I and II research of drug trials, and an inadequate 'dose' of acupuncture can bias results towards a type II error (a false-negative finding).

Thus, in the included reviews, several acupuncture treatment adequacy-related variables were evaluated, and the impact of these adequacy-related variables on treatment outcomes was assessed, using subgroup analyses. For the first review, on low back pain (Chapter 7), objective classification cut-points for all acupuncture adequacy-related variables were pre-specified. These adequacy cut-points were established based on an appraisal of empirical research studies, practice patterns of acupuncture, and acupuncture text books. For example, for the number of treatment sessions variable, a cut-point of six or more treatments versus fewer than six treatments was pre-specified because an earlier systematic review showed that six or more treatments was associated with better effects of acupuncture.[58] For the second review, on osteoarthritis (Chapter 8), a more subjective approach was used. This more subjective method involved having two experienced acupuncturists independently classify each of four acupuncture adequacy-related variables as adequate, inadequate, or unclear. These treatment adequacy classifications were made by the acupuncturists based on their subjective assessment of the description of the relevant variable in the text of each RCT publication, while blinded to any of the RCTs' results. This subjective assessment method did not involve the use of clear-cut, objective, pre-specified cut-points for defining categories of the treatment adequacy-related variables. It relied instead on the assessors' subjective judgments for assessing each variable. These subjective judgments may have resulted in considerable inter-observer variation or inconsistency of assessments across trials. Because of the potential for lower variation and higher consistency when using more objective classifications, for the acupuncture for IVF review update (Chapter 9), objective, pre-specified categories were used for classifying the treatment adequacy-related variables, whenever possible. The only exception was for assessing the qualifications of the RCTs' acupuncture practitioners. Acupuncturists' qualifications are difficult to classify based on pre-specified objective or quantitative criteria because qualifications may depend

on a subjective assessment that considers several different variables, including the acupuncturists' credentials, experience, and hands-on proficiency.

With some exceptions,[59] most of the acupuncture RCTs reviewed in thesis chapters 7-9 did not report hiring only optimally qualified and skilled acupuncturists. The large multi-center German trials included in the osteoarthritis review,[45,60] for example, required only 140 hours of acupuncture training in order for the physician-acupuncturists to be eligible to participate. Would more extensively trained and experienced acupuncturists result in better patient outcomes? One study addressed this question by pooling the data from four large German pragmatic trials, and found that neither the duration of training nor the duration of experience of the physician-acupuncturists influenced patients' outcomes after acupuncture.[61] However, this issue is far from resolved, and RCTs that use optimally qualified and skilled practitioners may give acupuncture the best opportunity to prove itself under optimal conditions; however, the results of such RCTs may not be generalizable to everyday practice.

2.2.4. Subgroup analyses and evaluating an RCT according to the efficacy-effectiveness continuum.

One additional RCT study design-related component that might be useful to extract for subgroup analyses in systematic reviews, but one which has rarely been considered in the past, is the efficacy versus effectiveness aspects of the RCT.[62,63] 'Efficacy' refers to the extent to which a specific intervention is beneficial under carefully controlled conditions chosen to maximize the likelihood of observing an effect if it exists. In contrast, 'effectiveness' is a measure of the extent to which an intervention, when deployed in the field in routine circumstances, does what it is intended to do for a specific population. There are multiple RCT criteria that need to be considered in evaluating RCTs along the efficacy-effectiveness continuum (e.g., patient eligibility criteria, treatment flexibility, practitioners' expertise, practitioners' adherence to protocol).[62] Some RCTs may be considered more effectiveness RCTs, as judged by some of these criteria, and more efficacy RCTs, as judged by other of these criteria. In addition, an RCT often cannot be unambiguously labeled as an efficacy or effectiveness trial. For example, an acupuncture RCT that uses both a sham acupuncture arm and a usual care arm, and a more flexible acupuncture intervention, might be judged to be in the middle of the efficacy-effectiveness continuum.[63] For the low back pain and osteoarthritis reviews (Chapters 7 and 8), we did not score the RCTs on the effectiveness-efficacy continuum. However, for the acupuncture as an adjuvant to IVF review (Chapter 9), we added an additional subgroup analysis based on whether the RCTs were conducted to test the effects of adjuvant acupuncture under controlled conditions in which the acupuncture was administered onsite at the IVF clinic versus RCTs conducted to test the effects of adjuvant acupuncture delivered offsite, which might better approximate every day, "real life" conditions since most IVF clinics do not have onsite acupuncturists. An instrument is under development to evaluate RCTs of acupuncture (and other nonpharmacological interventions), according to where they lie on the efficacy-effectiveness continuum.[63] These efficacy-effectiveness evaluations might be incorporated in future systematic reviews, for example by using subgroup analyses.

2.2.5. Selecting variables for the subgroup analyses.

As described above, for the systematic reviews that comprise this thesis, multiple variables related to the methodological quality and treatment adequacy of the included RCTs were extracted. Some of these variables are required for Cochrane reviews (i.e., the Cochrane risk of bias domains). Other variables (i.e., details of the treatment protocols) may be useful for practitioners, who may want to assess the relevance or feasibility of treatments for their own setting, or emulate the treatments tested in the RCTs with positive results. In addition, many of the variables extracted may be useful

to researchers conducting methodological studies examining the associations of RCT characteristics with intervention effect estimates using a collection of meta-analyses (i.e., meta-epidemiological studies). Therefore, it seems that there are few disadvantages in comprehensive data extraction, particularly for the electronically published Cochrane reviews, other than the additional time and effort that comprehensive data extraction requires. But which of these variables should be selected for subgroup analyses in the systematic reviews?

In the acupuncture for low back pain and acupuncture for osteoarthritis reviews (Chapters 7 and 8), the 'kitchen sink approach' was used, and subgroup analyses were conducted on all the variables extracted related to methodological quality. A disadvantage of this approach is that conducting too many subgroup analyses increases the risk of a spurious finding (a false positive finding). Therefore, for the acupuncture as adjuvant to IVF review update (Chapter 9), greater prudence was used in pre-specifying, in the protocol,[41] subgroup testing on only those methodological variables that would be expected to affect the estimates of the success of acupuncture on pregnancy, based on either empirical research or theoretical justification. For example, data on outcome assessor's blinding was extracted for this IVF review update, as required as an element of the Cochrane risk of bias tool. However, this variable was not selected for subgroup testing because it was deemed unlikely that the outcome assessors' knowledge of the treatment assignment (i.e., acupuncture versus control) would affect their assessment of whether or not the patient achieved a clinical pregnancy, ongoing pregnancy, or live birth, because these reproductive outcomes are entirely objective and unambiguous.[54]

In addition, only a few clinical characteristics were selected for subgroup testing for the acupuncture for IVF review, both because the relevant data for subgroup testing were not available for data extraction, and/or the subgroup variable was deemed unlikely to modify the effect of acupuncture on IVF success. That is, variables related to patient characteristics (e.g., age, duration of infertility) were not selected for subgroup testing because IVF RCT publications typically include heterogeneous populations and do not systematically report the effects of acupuncture across different categories of patient characteristics. Some of these patient-level variables, such as patient age, are prognostic of IVF success.[64] However, prognostic factors for IVF success are not necessarily good candidates for subgroup analyses unless such factors are also believed to modify the effect of acupuncture on IVF success.[56]

2.2.6. Correlation between subgroup variables.

A major challenge in interpreting the results of subgroup analyses for the comparison of true acupuncture versus sham acupuncture was the positive correlation between an increased intensity/complexity of the true acupuncture treatment and an increased risk of patient unblinding to treatment assignment. That is, treatment intensity and risk of unblinding are positively correlated because a more intensive/complex true acupuncture is associated with an increased risk of unblinding. And both increased treatment intensity and inadequate blinding would each independently predict greater reported benefits of true acupuncture relative to sham. Thus, if a systematic review finds that those sham-controlled RCTs that found a greater reported benefit of true acupuncture relative to sham acupuncture were the same RCTs that used more intensive true acupuncture treatments, there can be two potential explanations for this finding: 1) a more intensive, and thereby assumedly better, true acupuncture treatment was truly more effective in reducing RCT participants' symptoms or 2) in sham-controlled RCTs that used more intensive/complex true acupuncture treatments, unblinding of participants to treatment assignment (i.e., true or sham acupuncture) was more likely, and outcomes assessor bias due to incomplete blinding explains the greater reported benefit of true acupuncture relative to sham acupuncture.

An example from the Cochrane review of acupuncture for peripheral joint osteoarthritis (Chapter 8) illustrates how sham-controlled RCTs that used more intensive true acupuncture treatments were the same RCTs that were judged less likely to adequately maintain blinding. This made it impossible to disentangle whether the more intensive, and assumedly better, true acupuncture treatment, or alternatively, whether incomplete blinding due to this intensive true acupuncture treatment, was the causal factor in explaining the positive results. Namely, a subgroup analysis in the osteoarthritis review found that RCTs that used electro-acupuncture (i.e., a procedure in which the needles are electrically stimulated, which promotes stronger analgesic effects than manual stimulation of the needles alone, according to animal research[65]) were associated with better outcomes than RCTs in which the needles were not electrically stimulated, and the two sham-controlled RCTs in this review that used intensive electrical stimulation of all local knee points showed the greatest benefits.[66,67] These findings might indicate a superiority of electro-acupuncture over needle acupuncture without electrical stimulation. Alternatively, these findings may be explained by the fact that electro-acupuncture is probably more difficult to blind than needle acupuncture. As a result, some of the extra benefit seen in the electro-acupuncture RCTs may be due to incomplete blinding, leading to larger between group differences in placebo effects in electro-acupuncture RCTs than in RCTs in which the acupuncture needles were not electrically stimulated. It could not be conclusively determined whether or not the two sham-controlled RCTs that used intensive electro-acupuncture at all local knee points were adequately blinded, because neither tested intervention credibility. However, it seems reasonable to suggest that if the treatment group in an RCT received true electro-acupuncture (i.e., an intensive stimulation which is easily felt) and the control group received mock electro-acupuncture in which no electrical current is delivered, the additional co-intervention of the electrical current in only the true electro-acupuncture group would make it more likely for the RCT participants to figure out whether they were receiving a true or a sham treatment, than if the electrical current co-intervention were not used.

Another example illustrates how an increased true acupuncture treatment intensity or complexity may be more difficult to blind, and consequently more difficult to evaluate using sham-controlled RCTs. In the three reviews, the included RCTs were classified as using formula acupuncture (i.e., a simpler, standardised approach), or individualized acupuncture (i.e., a more complex treatment approach). In real-world practice, acupuncture treatment is often individualized, and continually modified to take into account changes in the patient's condition. However, individualization requires increased contact between the patient and the acupuncturist, which, again, increases the risk of unblinding. As it turned out, this individualization of treatment variable could not be analyzed in any of these reviews because of a dearth of RCTs that used individualized acupuncture. However, if it was found that sham-controlled RCTs that used individualized acupuncture had better reported outcomes than sham-controlled RCTs that used formula acupuncture, the causal explanation for this association may not have been possible to establish. That is, it could be explained either by the individualization of treatment being a more potent treatment than a formulaic treatment, or alternatively by a sub-optimal blinding in the individualized treatment sham-controlled RCTs.

2.3. Sensitivity Analyses

While subgroup analyses are used to compare the effects of treatment across subsets of RCTs, sensitivity analyses are used to explore the extent to which a review's findings are robust to the methodological decisions made in the process of conducting the review. Although there were several sensitivity analyses in the three reviews, the sensitivity analysis that was deemed most consequential, and that also generated a Letter to the Editor,[68] was the selection of the statistical

method for calculating SMDs (i.e., using changes from baseline values versus post-treatment values) for the primary analysis versus the sensitivity analysis. Namely, for the low back pain review (Chapter 7), only post-treatment values were pre-specified for calculating SMDs, and change from baseline values were not extracted or analyzed. A comparison of post-treatment values was pre-specified because this method may have an advantage over a comparison of changes from baseline values among trials in which standard deviations of changes from baseline are incompletely reported.[56,69] As was anticipated, this was the case for the older, incompletely reported RCTs included in this low back pain review. However, after the publication of this review in 2005, a Letter to the Editor[68] called into question the validity of this review's overall conclusions, arguing that the abstract conclusions were overstated, largely by pointing out that the SMD of one of the four individual sham-controlled RCTs in the analysis would have been smaller if the change from baseline values had been used instead of the post-treatment values. In response to this feedback, the SMDs for all RCTs were recalculated using change from baseline values, imputing standard deviations of change values, as necessary, and the results were not substantially changed. In subsequent reviews, both change from baseline outcome values (for the primary analyses) and post-treatment values (for the sensitivity analyses) were extracted (or obtained from RCT authors), in order to test whether the meta-analytic findings were robust to which of these two sets of outcome values were used.

2.4. Dealing with Unpublished RCTs

Unpublished RCTs is a catch-all term that could refer to completed but *never* published RCTs, ongoing RCTs, or completed but *not yet* published RCTs. Identifying completed but *never* published and ongoing acupuncture RCTs may remain difficult if acupuncture RCTs are not adequately 'registered' on databases of ongoing trials. This may indeed be the case, as *none* of the nine new RCTs that were subsequently published and included in the *updated* version of the adjuvant acupuncture for IVF review (Chapter 9) were included in ongoing trials databases at the time these databases were searched during 2007, for the *initial BMJ* version[70] of this review. With a lack of trial registration, it is very difficult for systematic review authors in CAM to know about RCTs in the pipeline, and therefore review authors cannot estimate how likely the evidence base is to change in the near future with the publication of new RCTs. More importantly, it raises the possibility that the published acupuncture/CAM RCTs are an incomplete, and perhaps unrepresentative, sample of all the relevant RCTs that have been conducted. This issue of a possible "publication bias" was discussed in the osteoarthritis and IVF reviews (Chapters 8 and 9). A possible future study to address this issue is to survey RCT authors in CAM to better understand the barriers to their registering their RCTs. However, this issue may likely solve itself in time, as prior RCT registration is increasingly becoming a hard requirement for journal publication.[71,72]

The issue of ongoing and completed but not *yet* published RCTs was probably a more relevant concern in these three reviews, particularly in drawing conclusions about efficacy. Namely, for the acupuncture for low back pain review (Chapter 7), one of the peer-reviewers suggested caution in our conclusions primarily because of the existence of four large, recently completed RCTs (two of which included a sham control), which, once published, would have a major impact on the overall evidence picture. The potential impact of these RCTs was addressed in the concluding paragraph of this low back pain review's Discussion section, where it was mentioned that the current evidence should be considered "still somewhat preliminary" because of the existence of these recently completed RCTs. However, the review's abstract conclusions statement did not adequately reflect the potential impact of these large and rigorous recently completed RCTs, which, once published, could, and indeed did (as later turned out to be the case), cause a major shift in our meta-analysis effect estimates. Therefore a correction of our abstract conclusions was published.[7] The corrected

conclusions better represented the preliminary nature of the evidence from the four small sham-controlled RCTs included in our review, and the potential effect of the large and rigorous recently completed RCTs on our preliminary estimates.

The broader question being addressed here is whether systematic review authors should draw bottom line conclusions based strictly on the evidence that they have reviewed, or whether they should also consider how recently completed, ongoing, or even future RCTs might impact this evidence. One consideration in making this decision is that the credibility of major medical journals is based largely on the trustworthiness of the studies they publish. Major journals would also likely prefer to publish systematic reviews with robust findings that can guide treatment decision-making for many years to come, and that will not be changed or refuted with the imminent publication of new RCTs. In light of this, and based on the lessons learned from the overstated abstract conclusions in the acupuncture for low back pain review, greater caution was used in drafting conclusions for the subsequent reviews, taking into consideration the strengths and limitations of the included RCTs, and also considering how the ongoing and recently completed RCTs, when published, might impact the existing evidence. This strategy may have been prudent, in that the abstract conclusions for the acupuncture for osteoarthritis and acupuncture for IVF reviews (Chapters 8 and 9) were sufficiently cautious to be robust even in light of RCTs published subsequent to these reviews that failed to show a benefit of acupuncture. At the same time, the abstract conclusions well reflected the RCTs included and analyzed in these reviews.

The stability of abstract conclusions to recently completed, but not yet published, RCTs that might affect these conclusions is probably less of a concern for Cochrane reviews than for print journal reviews, because Cochrane reviews are published electronically and are regularly amended and updated (although there is still "room for improvement"[73] in the frequency of this updating). Thus, for example, shortly after our Cochrane review of acupuncture for peripheral joint osteoarthritis (Chapter 8) was published in January 2010, a large new sham-controlled RCT was published in September 2010,[74] the inclusion of which may have affected our Cochrane review's overall meta-analytic effect estimates. Therefore, our 2010 Cochrane review was "withdrawn" by the Cochrane Musculoskeletal Review Group until it could be updated to include evidence from this recently published RCT.[74] In contrast, reviews in print journals are not updated, and cannot be withdrawn after publication because of a change in the evidence landscape, and therefore the conclusions of reviews in major print journals may need to be more robust to evidence that may be published in the near future.

2.5. Clinical Relevance

There are several factors that must be considered when deciding whether or not to include an intervention as part of standard medical care. Namely, data on the magnitude of the clinical benefits of an intervention need to be considered together with data on potential adverse reactions, patient preferences, and costs. In addition, the relative benefits of an intervention under consideration might be compared to the estimated relative benefits of alternative interventions that might be offered, or that are currently offered. Informed patients, in a partnership with their clinicians, must weigh all of these factors in deciding whether estimates of benefit from an intervention are clinically important.

2.5.1. Comparing meta-analysis effect estimates with established clinical relevance thresholds: uses and limitations.

An important determination in deciding whether an intervention is clinically useful is evaluating whether the effect sizes in the meta-analyses meet effect size thresholds for clinical relevance, or "minimal clinically important differences" (MCIDs). There is much debate about the meaning and

interpretation of MCIDs.[75,76] There is also a lack of consensus on the definition of MCIDs, and a variety of different methods have been used to estimate MCIDs.[75,76] For example, some authors define a change on a global rating scale of "minimal improvement" or "slightly better" as the MCID threshold, whereas other authors would not classify such a small improvement as meeting the MCID threshold.[76]

Researchers have established such MCID thresholds for both low back pain and osteoarthritis. For the 2005 low back pain review (Chapter 7), the MCIDs used were those established by an expert consensus from the Cochrane Back Review Group (CBRG) Editorial Board.[77-79] For this low back pain review, the meta-analytic effect estimates for the comparison of acupuncture versus sham acupuncture exceeded the CBRG's thresholds for MCIDs; however, in the large sham-controlled RCTs published subsequent to this review, the effects of acupuncture versus sham did not meet these MCID thresholds for low back pain. For the osteoarthritis review (Chapter 8), the MCID thresholds used were those generated by a prospective cohort study that estimated the within group changes on the WOMAC score that corresponded to a self-perceived change of "slightly better" compared to baseline.[80] For this osteoarthritis review, the results of the acupuncture versus sham comparison did not meet these MCID thresholds. Therefore for the osteoarthritis review, the conclusion was that the effects of acupuncture relative to sham were too small to be perceived by patients as beneficial, and would therefore be considered clinically meaningless or irrelevant for individual patients.[81] However, an intervention effect that may be considered small for individuals may still be important, from a societal point of view, for conditions with a high prevalence, such as osteoarthritis and low back pain (i.e., by causing a small shift in the populations affected by these conditions, in terms of work absences, healthcare use, prevention of long-term disability). MCIDs are more typically applied in the context of pragmatic RCTs, in terms of helping decision-makers interpret to what extent overall the effect of the test treatment, relative to no treatment or an alternative treatment, is worthwhile. In both the low back pain and osteoarthritis reviews, the effects of acupuncture relative to a waiting list control, and some of the other active treatment control groups, did exceed the pre-specified MCID thresholds for clinical relevance.

In determining the potential clinical usefulness of a treatment, and in interpreting clinical relevance thresholds, another issue to consider may be the alternative treatments that are available for the given condition and the magnitude of effect of these alternative treatments. For example, while in the osteoarthritis review (Chapter 8) the effects of acupuncture relative to sham acupuncture (SMD -0.28, 95% CI -0.45 to -0.11) did not meet the pre-specified clinical relevance thresholds, few, if any, other commonly used pharmacological or nonpharmacological treatments for osteoarthritis meet these thresholds[80] for minimal clinically important differences. For example, NSAIDS (relative to an inert placebo) do not meet these thresholds,[80,82] yet NSAIDs are used regularly by one-third of all people with osteoarthritis.[83] Acetaminophen, also one of the most widely used medications for osteoarthritis,[83,84] has effects, relative to placebo, far below this threshold (SMD -0.13, 95% CI -0.22 to -0.04).[85] In addition, acetaminophen can cause serious liver damage, even when taken in small amounts more than the recommended dose.[86] Finally, intra-articular injection of hyaluronic acid is also widely used for knee osteoarthritis, but a 2012 systematic review[87] concluded that this intervention "is associated with a small and clinically irrelevant benefit and an increased risk for serious adverse events." However, some caution is warranted in comparing, for example, the effect sizes of NSAIDS (or acetaminophen or hyaluronic acid) relative to placebo pills versus the effect sizes of acupuncture relative to sham acupuncture because this involves a non-randomized comparison, and also because the control interventions differ (e.g., placebo pills versus sham acupuncture). Nevertheless, it is still true that few if any osteoarthritis treatments have specific effects that meet the thresholds for clinically relevant benefits. Therefore one might question whether the effects of acupuncture should be required to

meet this threshold, which other osteoarthritis therapies also do not meet. Indeed, it may be that the threshold for clinical relevance is too high for any individual treatment alone, and that a multidisciplinary approach to treating osteoarthritis, with a focus on combining several therapies is necessary. In fact, in a recent RCT using such a multidisciplinary treatment approach, although for treating fibromyalgia, the effect sizes were larger than the MCID thresholds, for a substantial proportion of RCT participants.[88]

In the acupuncture for osteoarthritis review (Chapter 8), as well as in the NSAIDs[82] and hyaluronic acid[87] reviews mentioned above, the clinical relevance MCID thresholds used were derived from a prospective cohort study examining within person changes over time *in a single group*. However, these MCID estimates were then applied to determine whether differences in changes *between the two groups in an RCT* were meaningful. A major limitation of this application is that these MCIDs used, derived from changes over time within a single group, are likely to be greater than those that would have been estimated from differences in changes between a test and a control group over time.[75] This is because the control group in an RCT would also be expected to show some improvement over time. Thus, although these MCIDs were used to help with the interpretation of differences in outcomes between the acupuncture and control groups in our osteoarthritis review, it may be difficult for any test treatment to meet these high MCIDs, when a test treatment is compared to a control that is also likely to show improvement. In addition, an MCID may be different for every RCT, unless exactly the same treatments are compared in the same populations, and in the same settings. Finally, MCID thresholds should not be considered immutable and fixed values to be used in all cases, but instead they may be patient or context dependent. For example, a smaller MCID threshold might be used for a simple, cheap, and safe intervention such as acupuncture, whereas a larger MCID may be more appropriate for a more expensive or risky procedure, such as surgery.[76]

2.5.2. Weighing benefits versus adverse effects in establishing acupuncture's clinical usefulness.

In addition to evaluating the estimated effect sizes of acupuncture, and whether or not the estimated benefits are clinically relevant, another important consideration in deciding whether acupuncture should be included as part of standard care is deciding whether any benefits outweigh any risks and costs. Incidence rates for major adverse effects of acupuncture are best estimated from large prospective surveys of practitioners. Four such surveys of acupuncture safety, two conducted in Germany,[89,90] and two in the United Kingdom,[91,92] confirm that serious adverse events after acupuncture are uncommon.

In addition to assessing and summarizing these previously published prospective surveys evaluating acupuncture's adverse effects, data on adverse events were also extracted from the RCTs included in the reviews, starting with the osteoarthritis review, and continuing with the IVF review. For the osteoarthritis review (Chapter 8), the frequency of adverse events was found to be similar between the acupuncture and control groups. However, pooling of adverse events across these osteoarthritis RCTs was not possible because of limited and unclear reporting, and heterogeneous definitions of adverse events. For the acupuncture as an adjuvant to IVF review (Chapter 9), the primary adverse event that was evaluated was miscarriage rate. However, data on other potential adverse events of acupuncture or control (i.e., pain, anxiety) were also extracted because it was judged that the potential for adverse events is of primary importance in deciding whether the low-cost and potentially effective adjuvant acupuncture for IVF is clinically useful.

2.5.3. Considering cost-effectiveness in establishing acupuncture's clinical usefulness.
Finally, because government health agencies and other funders of health care have limited resources, another important consideration in determining whether or not to include an intervention as part of standard care is evaluating its benefits relative to its costs. The benefit-cost ratio of acupuncture, as evaluated using cost-effectiveness studies, may be among the most important factors to consider in determining acupuncture's clinical usefulness. This is because other research has already established that acupuncture has a low risk of adverse effects[89-92] and is accepted by patients.[93] However, it might be argued that acupuncture should first demonstrate specific benefits relative to sham acupuncture before it is evaluated for its cost-effectiveness. While the reviews included in this thesis did not definitively conclude that acupuncture is more effective than sham acupuncture for low back pain or osteoarthritis, the most current, comprehensive, and authoritative review to date of acupuncture for pain-related conditions (i.e., a 2012 individual patient data (IPD) meta-analysis of 29 RCTs (n=17, 922)[94]) concluded that there were small but robust differences between true and sham acupuncture that could be clearly distinguished from bias, for each of the four pain-related conditions evaluated, including low back pain and osteoarthritis. There have been no published letters to the Editor that have challenged the methodology or conclusions of this IPD meta-analysis. This IPD therefore provides support that acupuncture has specific effects relative to sham acupuncture, and consequently provides support for cost-effectiveness evaluations of acupuncture.

Such cost-effectiveness analyses are typically applied to pragmatic RCTs that compare interventions as they are offered in clinical practice, to support decision making and guideline recommendations. On the other hand, cost-effectiveness analyses are not usually applied to sham-controlled RCTs because sham acupuncture is not an intervention offered in clinical practice. While cost-effectiveness was not considered as an outcome in the included reviews, there have been two cost-effectiveness RCT publications for low back pain[95,96] and three for osteoarthritis[97-99] (one of which was published as an abstract[97]). The cost-effectiveness outcome used in all five of these RCTs[95-99] was the incremental cost incurred per quality adjusted life year gained. All five of these cost-effectiveness analyses found that acupuncture treatment was associated with significantly higher costs compared to usual care, with this increase in costs mostly due to the costs of the acupuncture. However, these cost-effectiveness analyses also all found that acupuncture provided a cost-effective use of health care resources despite an associated increase in costs. That is, in all of these analyses,[95-99] the incremental cost-effectiveness ratio per quality-adjusted life year gained met UK threshold values for a cost-effective intervention, which has been defined as 20,000UK£ to 30,000UK£ (corresponding to approximately $25,000 to $37,000 USD) per quality adjusted life year gained.[100] While these RCTs have found that acupuncture is a cost-effective intervention, it is not clear whether this favorable cost-benefit ratio is a result of a specific effect of acupuncture needling, or nonspecific effects (e.g., greater autonomy and empowerment of patients, positive patient–practitioner relationship).[98]

2.6. Systematic Reviews of Acupuncture: Implications for Future Research
This final section summarizes the questions raised by the three systematic reviews included in this thesis, and provides recommendations on future important questions to address, and the optimal study designs for addressing these questions.

One important future research question is whether different subgroups of participants benefit differently from acupuncture. That is, the 'summary data' systematic reviews included in this thesis could only address whether *trial-level variables* (e.g., type of sham; number of acupuncture treatments) can affect the success of acupuncture. However, to address the question of whether *individual patient-level variables* (e.g., age) affect acupuncture treatment success, individual patient

data (IPD) systematic reviews are necessary. The application of the IPD methodology to the review of the topic area of acupuncture as an adjuvant to IVF is being tentatively considered, to address whether specific IVF subgroups may benefit more from adjuvant acupuncture. That is, a subgroup analysis in the summary data review of this topic (Chapter 9), found that the addition of acupuncture increased the pregnancy success rates in RCTs with lower baseline (i.e., control group) IVF pregnancy rates. However, adjuvant acupuncture did not further increase pregnancy success rates in RCTs with already high IVF baseline pregnancy rates. Because baseline pregnancy rate is possibly a proxy for the number of embryos transferred (although other factors are also involved[64]), an IPD meta-analysis may be helpful to further investigate the relationship between the number of embryos transferred, as well as other individual patient-level variables (e.g., age), and the efficacy of adjuvant acupuncture. This question can only be addressed using the IPD approach.

The remaining important research questions can largely be addressed only through new RCTs with innovative designs. Namely, such RCTs, and the subsequent systematic reviews of these RCTs, will provide a better understanding of whether acupuncture is more effective than a sham acupuncture intervention lacking acupuncture-specific effects, and also the relative effects of acupuncture compared to other active treatments.

For the 2010 osteoarthritis review (Chapter 8), no firm conclusions could be drawn about whether acupuncture is more effective than a sham acupuncture intervention that is believable, but that also lacks acupuncture-specific effects. This was because all the sham control interventions that were judged as believable to RCT participants as true acupuncture were also all judged to have acupuncture-specific effects. The pooled results from these RCTs did not show a clear benefit of true acupuncture relative to sham acupuncture on the outcomes (e.g., pain). However, this may have been because the sham acupuncture interventions decreased pain through the same putative mechanism as the true acupuncture intervention. Therefore, the research recommendation for future RCTs is to use a convincing sham acupuncture intervention that is as believable to RCT participants as true acupuncture, but that will not affect the outcome through the same mechanisms as that of the putative pathway of true acupuncture. Perhaps the best way to accomplish this is to use non-penetrating, but demonstrably believable, sham needles that are placed far away from the true acupuncture points. In addition, such sham-controlled RCTs should ideally be restricted to acupuncture naïve participants because such participants would be less likely to be able to differentiate between the needling from the true acupuncture versus the non-penetrating sham. While this recommendation relates to the design of future sham-controlled RCTs, it might be argued that the acupuncture versus sham acupuncture comparison has already been given too much emphasis in previous RCTs; that sham-controlled RCTs can tell only part of the story of the effectiveness of acupuncture; and that future RCTs should instead shift from sham controls to active controls. Namely, the emphasis on whether or not acupuncture has small, clinically irrelevant specific benefits compared to sham acupuncture might distract from the primary 'selling points' of acupuncture, which are that it is inexpensive, has few side effects, and has a potent placebo effect. Indeed, because acupuncture may have such a potent placebo effect, which may be valuable in empowering patients, researchers might design future RCTs that maximize, rather than minimize, these placebo effects (as further described below), by focusing less on measuring the specific effects of acupuncture relative to sham, a design which controls for, and thereby minimizes, these placebo effects. However, on the other hand, sham-controlled RCTs are also highly informative and valuable because a demonstration that true acupuncture has a specific benefit over sham acupuncture may be important for physicians and third-party payers, who might be reluctant to refer patients to, or reimburse patients for, a treatment which is strictly a placebo, no matter how potent. Finally, as discussed in section 2.2.1.2, the value of a sham control may depend on the research question of interest and the outcome being evaluated. Namely, there are two completely

different reasons to use a sham control in acupuncture RCTs: 1) to control for biased patient outcome reporting, and 2) to tease out the specific effects of a treatment. Using a sham control to blind patients to get unbiased outcomes reports seems justified and reasonable for acupuncture RCTs evaluating subjective, patient-reported outcomes (e.g., pain). However, using a sham control in RCTs with entirely objective outcomes (e.g., clinical pregnancy) may only be justified if the objective of the RCT is to understand the mechanism of acupuncture's effects.

Another research recommendation is to conduct pragmatic RCTs to investigate the effects of the entire 'package' of acupuncture, as offered in regular clinical practice by acupuncturists. That is, in clinical practice, acupuncture is often not used as a singular and fixed treatment, but rather combined with other traditional Chinese medicine treatments. Namely, acupuncturists may treat patients using individualized acupuncture, while also including Chinese herbal medicine, tai chi, and diet and lifestyle recommendations. Trials that evaluate such a multi-modal and flexible approach may better represent the treatment patients would receive by visiting a traditional acupuncturist. But what would be an appropriate control for such a trial design? One possibility may be to compare the entire 'package' of acupuncture versus an alternative complete, complex system of care (e.g., multi-modal medical care; physical therapy). Indeed, there have already been RCTs that have included a comparison of acupuncture alone versus multi-modal standard medical care. Such RCTs have typically found that acupuncture produced better results than standard medical care on the patient reported subjective outcome measures (e.g., patient reported pain and functional status). However, the interpretation of these results is often complicated by the fact that sham acupuncture, though found equivalent to true acupuncture, was also found better than standard medical care. Based on these difficult to interpret findings, some commentators[101] have argued that the better patient-reported outcomes in the true *and* sham acupuncture treatment groups, compared to the standard medical care group, can be entirely explained by the higher expectation effects of acupuncture relative to standard medical treatment. Indeed, because acupuncture may elicit a greater expectation effect than standard medical care,[4,101,102] particularly among participants who have a preference for acupuncture,[103] investigators conducting future pragmatic RCTs that compare acupuncture with other active therapies might consider asking participants about their expectations and preferences (before and after the intervention). Such trials might also include a credibility questionnaire to establish that the acupuncture and the active treatment it is being compared with are perceived by the RCT participants as equally credible treatments.[103] However, on the other hand, it might be argued that acupuncture's placebo effects are 'part and parcel' of the acupuncture treatment, and that there is no need to evaluate, or attempt to equalize, placebo effects between treatments because pragmatic RCTs should be used to best approximate the average likely response to treatment in clinical practice, in which treatment effects and placebo factors, expectation effects, and patient preferences may all operate.[104] Finally, the fact that *sham acupuncture,* which is ostensibly a placebo treatment, has typically outperformed standard medical care suggests that medical doctors underperform, in the sense that they forget to let the placebo effect do most of the work.

Conclusions

The likely potent placebo effects of acupuncture have presented the major methodological challenge in the interpretation of acupuncture trials and reviews. That is, acupuncture seems to have only small benefits relative to placebo/sham controls, but often clear, clinically relevant benefits compared to usual care controls, according to the most current, comprehensive, and authoritative IPD review to date.[94] Which RCT design should policy decisions be based upon? That is, should RCTs attempt to control for and minimize acupuncture's placebo effects using the sham-controlled design, or instead, allow for and maximize acupuncture's potent placebo effects, using the

comparative effectiveness design? While most early Western RCTs used the sham-controlled RCT design to estimate the point specific effects of needle placement, RCTs are now increasingly comparing acupuncture to other active treatments. Trials published in China, in particular, often compare acupuncture with other active treatments, and rarely to a sham control. The lack of blinding is the major problem in interpreting the results of these comparative effectiveness trials, particularly because the outcome measures evaluated in these trials are typically subjective and patient reported, and acupuncture may have more potent placebo effects than other treatments.[52,102] Indeed, Chinese comparative effectiveness trials in particular almost always find acupuncture superior to the active control treatment.[16] As more and more of these Chinese trials are becoming available to systematic review authors through the Cochrane Collaboration's CAM Field trials database, major challenges will be whether and how to incorporate these trials into systematic reviews, and how to interpret the resulting reviews, e.g., for the development of guidelines. For example, a 2012 Cochrane review of acupuncture for irritable bowel syndrome (IBS)[103] found that acupuncture is equivalent to sham acupuncture, based on Western RCTs, but superior to commonly used pharmacological treatments, based on Chinese RCTs. However, the superiority of acupuncture relative to pharmacological therapy in these Chinese IBS RCTs may have been entirely due to the larger placebo effects of acupuncture relative to drugs. Based on such findings, should guideline developers recommend acupuncture, or not? Such decisions should perhaps await future RCTs that will likely clarify whether or not these reportedly greater benefits of acupuncture relative to other active treatments are due entirely to participants' preferences for acupuncture or participants' greater expectations of improvement on acupuncture relative to other more commonly used therapies.[103] After such additional, well-designed RCTs have been completed, it will be easier for systematic review authors to provide valid and reliable estimates of the effects of acupuncture for osteoarthritis, low back pain, and IVF, being informed by several different levels of evidence, each with its own complexities and inherent biases. As Cochrane CAM-related reviews are updated to include such RCTs, the Cochrane CAM Field topics list will be an increasingly valuable resource for patients, practitioners, and policy-makers.

REFERENCES

1. Manheimer E, Wieland S, Kimbrough E, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for Traditional Chinese Medicine therapies. *J Altern Complement Med* 2009;15(9):1001-14.
2. Wieland L*, Manheimer E*, Berman B. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane Collaboration. *Altern Ther Health Med* 2011;17(2):50-9. *Both authors contributed equally to this work.
3. Wieland LS, Manheimer E, Sampson M, Barnabas JP, Bouter LM, Cho K, Lee MS, Li X, Liu J, Moher D, et al. Bibliometric and content analysis of the Cochrane Complementary Medicine Field specialized register of controlled trials. *Syst Rev* 2013 July 4;2:51.
4. Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, van der Windt DA, Berman BM, Bouter LM. Acupuncture for peripheral joint osteoarthritis. *Cochrane Database Syst Rev* 2010(1):CD001977.
5. Manheimer E*, Ezzo J*. Chapter 2: Systematic reviews and CAM. In: Adams J, editor. *Researching Complementary and Alternative Medicine*. Oxford: Routledge, 2006:19-36. *Both authors contributed equally to this work.
6. Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. *J Clin Epidemiol* 2006;59(2):107-13.
7. Manheimer E, White A, Berman B, Forys K, Ernst E. Meta-analysis: acupuncture for low back pain. *Ann Intern Med* 2005;142(8):651-63. Erratum in: *Ann Intern Med*. 2005 Jun 7;142(11):950-1.
8. Manheimer E, van der Windt D, Cheng K, Stafford K, Liu J, Tierney J, Lao L, Berman BM, Langenberg P, Bouter LM. The effects of acupuncture on rates of clinical pregnancy among women undergoing in vitro fertilization: a systematic review and meta-analysis. *Hum Reprod Update* 2013 June 27; doi:10.1093/humupd/dmt026. [Epub ahead of print]
9. Angell M, Kassirer JP. Alternative medicine--the risks of untested and unregulated remedies. *N Engl J Med* 1998;339(12):839-41.
10. Bland M. Re: Assaulting alternative medicine: worthwhile or witch hunt. *BMJ* 2012;344:e1075.
11. Cassidy CM. Commentary on terminology and therapeutic principles: challenges in classifying complementary and alternative medicine practices. *J Altern Complement Med* 2002;8(6):893-5.
12. Turner LA, Singh K, Garritty C, Tsertsvadze A, Manheimer E, Wieland LS, Galipeau J, Moher D. An evaluation of the completeness of safety reporting in reports of complementary and alternative medicine trials. *BMC Complement Altern Med* 2011;11:67.
13. Herman PM, Poindexter BL, Witt CM, Eisenberg DM. Are complementary therapies and integrative care cost-effective? A systematic review of economic evaluations. *BMJ Open* 2012;2(5).
14. Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials* 2009;10:46.
15. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7(1):1-76.
16. Vickers A, GoyalN, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;19(2):159-66.
17. Yoke YK, Price D, Herxheimer A (editors). Chapter 14: Adverse effects. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (Accessed 2013 July 12).
18. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med* 2011;8(5):e1001026.
19. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;18(4):476-99.
20. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;16(2):137-62.
21. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10(1):31-56.
22. Jorgensen AW, Hilden J, Gotzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ* 2006;333(7572):782.
23. Olsen O, Middleton P, Ezzo J, Gotzsche PC, Hadhazy V, Herxheimer A, Kleijnen J, McIntosh H. Quality of Cochrane reviews: assessment of sample from 1998. *BMJ* 2001;323(7317):829-32.

24. Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions (section 12.2.2). In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (Accessed 2013 July 12).
25. Lasserson T. Audit of abstracts and summaries. *Cochrane Editorial Unit Bulletin* 2011 August;1. Available from http://www.editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/Abstract%20audit%20report.pdf (Accessed 2013 July 12).
26. Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, Dilkes H. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. *Med Decis Making* 2010;30(5):566-77.
27. Cates CJ, Oleszczuk M, Stovold E, Wieland LS. Safety of regular formoterol or salmeterol in children with asthma: an overview of Cochrane reviews. *Cochrane Database Syst Rev* 2012;10:CD010005.
28. Jones L, Othman M, Dowswell T, Alfirevic Z, Gates S, Newburn M, Jordan S, Lavender T, Neilson JP. Pain management for women in labour: an overview of systematic reviews. *Cochrane Database Syst Rev* 2012;3:CD009234.
29. Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database Syst Rev* 2011(7):CD009255.
30. Keus F, Gooszen HG, van Laarhoven CJ. Open, small-incision, or laparoscopic cholecystectomy for patients with symptomatic cholecystolithiasis. An overview of Cochrane Hepato-Biliary Group reviews. *Cochrane Database Syst Rev* 2010(1):CD008318.
31. Ryan R, Santesso N, Hill S, Lowe D, Kaufman C, Grimshaw J. Consumer-oriented interventions for evidence-based prescribing and medicines use: an overview of systematic reviews. *Cochrane Database Syst Rev* 2011(5):CD007768.
32. Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, Tanjong Ghogomu E, Tugwell P. Biologics for rheumatoid arthritis: an overview of Cochrane reviews. *Cochrane Database Syst Rev* 2009(4):CD007848.
33. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet* 2002;359(9310):966-70.
34. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000;136(3):381-5.
35. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152(11):726-32.
36. MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, White A, Moher D. Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): Extending the CONSORT statement. *J Evid Based Med* 2010;3(3):140-55.
37. Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol* 2012;65(6):602-9.
38. Manheimer E. Selecting a control for in vitro fertilization and acupuncture randomized controlled trials (RCTs): how sham controls may unnecessarily complicate the RCT evidence base. *Fertil Steril* 2011;95(8):2456-61.
39. Langevin HM, Wayne PM, Macpherson H, Schnyer R, Milley RM, Napadow V, Lao L, Park J, Harris RE, Cohen M, et al. Paradoxes in acupuncture research: strategies for moving forward. *Evid Based Complement Alternat Med* 2011;2011:180805.
40. Ahn AC, Colbert AP, Anderson BJ, Martinsen OG, Hammerschlag R, Cina S, Wayne PM, Langevin HM. Electrical properties of acupuncture points and meridians: a systematic review. *Bioelectromagnetics* 2008;29(4):245-56.
41. Manheimer E, van der Windt DA, Stafford K, Liu J, Tierney J, Lao L, Berman B, Bouter L. Effects of adjuvant acupuncture on pregnancy rates among women undergoing in vitro fertilization: systematic review and meta-analysis. PROSPERO 2011:CRD42011001309. Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001309 (Accessed 2013 July 12).
42. Miller FG, Kaptchuk TJ. Acupuncture trials and informed consent. *J Med Ethics* 2007;33(1):43-4.
43. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;328(7437):432.
44. Linde K, Streng A, Jurgens S, Hoppe A, Brinkhaus B, Witt C, Wagenpfeil S, Pfaffenrath V, Hammes MG, Weidenhammer W, et al. Acupuncture for patients with migraine: a randomized controlled trial. *JAMA* 2005;293(17):2118-25.
45. Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, Hummelsberger J, Walther HU, Melchart D, Willich SN. Acupuncture in patients with osteoarthritis of the knee: a randomised trial. *Lancet* 2005;366(9480):136-43.
46. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol* 2007;36(3):664-5.
47. Altman DG, Schulz KF, Moher D. Turning a blind eye: testing the success of blinding and the CONSORT statement. *BMJ* 2004;328(7448):1135; author reply 36.

48. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;359(9307):696-700.
49. Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;328(7448):1136.
50. Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. *Lancet* 2010;375(9721):1144-6.
51. Hrobjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev* 2004(3):CD003974.
52. Wechsler ME, Kelley JM, Boyd IO, Dutile S, Marigowda G, Kirsch I, Israel E, Kaptchuk TJ. Active albuterol or placebo, sham acupuncture, or no intervention in asthma. *N Engl J Med* 2011;365(2):119-26.
53. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336(7644):601-5.
54. Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (Accessed 2013 July 12).
55. Mallett S, Clarke M. The typical Cochrane review. How many trials? How many participants? *Int J Technol Assess Health Care* 2002;18(4):820-3.
56. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (Accessed 2013 July 12).
57. van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine (Phila Pa 1976)* 2009;34(16):1685-92.
58. Ezzo J, Berman B, Hadhazy VA, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;86(3):217-25.
59. Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, Deyo RA. Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. *Arch Intern Med* 2001;161(8):1081-8.
60. Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN. Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm. *Arthritis Rheum* 2006;54(11):3485-93.
61. Witt CM, Ludtke R, Wegscheider K, Willich SN. Physician characteristics and variation in treatment outcomes: are better qualified and experienced physicians more successful in treating patients with chronic pain with acupuncture? *The journal of pain : official journal of the American Pain Society* 2010;11(5):431-5.
62. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ* 2009;180(10):E47-57.
63. Witt CM, Manheimer E, Hammerschlag R, Ludtke R, Lao L, Tunis SR, Berman BM. How well do randomized trials inform decision making: systematic review using comparative effectiveness research measures on acupuncture for back pain. *PLoS One* 2012;7(2):e32399.
64. van Loendersloot LL, van Wely M, Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;16(6):577-89.
65. Ulett GA, Han S, Han JS. Electroacupuncture: mechanisms and clinical application. *Biol Psychiatry* 1998;44(2):129-38.
66. Vas J, Mendez C, Perea-Milla E, Vega E, Panadero MD, Leon JM, Borge MA, Gaspar O, Sanchez-Rodriguez F, Aguilar I, et al. Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. *BMJ* 2004;329(7476):1216.
67. Sangdee C, Teekachunhatean S, Sananpanich K, Sugandhavesa N, Chiewchantanakit S, Pojchamarnwiputh S, Jayasvasti S. Electroacupuncture versus diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. *BMC Complement Altern Med* 2002;2:3.
68. Shekelle P. Acupuncture for low back pain. *Ann Intern Med* 2005;143(9):691-2; author reply 92-3.
69. Higgins JPT, Deeks JJ, Altman DG (editors). Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (Accessed 2013 July 12).
70. Manheimer E, Zhang G, Udoff L, Haramati A, Langenberg P, Berman BM, Bouter LM. Effects of acupuncture on rates of pregnancy and live birth among women undergoing in vitro fertilisation: systematic review and meta-analysis. *BMJ* 2008;336(7643):545-9.
71. Laine C, De Angelis C, Delamothe T, Drazen JM, Frizelle FA, Haug C, Hebert PC, Horton R, Kotzin S, Marusic A, et al. Clinical trial registration: looking back and moving ahead. *Ann Intern Med* 2007;147(4):275-7.
72. Rennie D. Trial registration: a great idea switches from ignored to irresistible. *JAMA* 2004;292(11):1359-62.
73. Kristiansen IS. How up-to-date are Cochrane reviews? *Lancet* 2008;371(9610):384; author reply 84-5.

74. Suarez-Almazor ME, Looney C, Liu Y, Cox V, Pietz K, Marcus DM, Street RL, Jr. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. *Arthritis Care Res (Hoboken)* 2010;62(9):1229-36.
75. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002;14(2):109-14.
76. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008;33(1):90-4.
77. Assendelft WJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG. Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Ann Intern Med* 2003;138(11):871-81.
78. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)* 2003;28(12):1290-9.
79. van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. *Spine (Phila Pa 1976)* 1997;22(20):2323-30.
80. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol* 2002;29(1):131-8.
81. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;45(4):384-91.
82. Bjordal JM, Ljunggren AE, Klovning A, Slordal L. Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: meta-analysis of randomised placebo controlled trials. *BMJ* 2004;329(7478):1317.
83. Ausiello JC, Stafford RS. Trends in medication use for osteoarthritis treatment. *J Rheumatol* 2002;29(5):999-1005.
84. Wegman A, van der Windt D, van Tulder M, Stalman W, de Vries T. Nonsteroidal antiinflammatory drugs or acetaminophen for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines. *J Rheumatol* 2004;31(2):344-54.
85. Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. *Cochrane Database Syst Rev* 2006(1):CD004257.
86. Kuehn BM. FDA focuses on drugs and liver damage: labeling and other changes for acetaminophen. *JAMA* 2009;302(4):369-71.
87. Rutjes AW, Juni P, da Costa BR, Trelle S, Nuesch E, Reichenbach S. Viscosupplementation for osteoarthritis of the knee: a systematic review and meta-analysis. *Ann Intern Med* 2012;157(3):180-91.
88. Castel A, Fontova R, Montull S, Perinan R, Poveda MJ, Miralles I, Cascon R, Hernandez P, Aragones N, Salvat I, et al. The efficacy of a multidisciplinary fibromyalgia treatment adapted for women with low educational levels: A randomised, controlled trial. *Arthritis Care Res (Hoboken)* 2012;65(3):421-31.
89. Melchart D, Weidenhammer W, Streng A, Reitmayr S, Hoppe A, Ernst E, Linde K. Prospective investigation of adverse effects of acupuncture in 97 733 patients. *Arch Intern Med* 2004;164(1):104-5.
90. Witt CM, Pach D, Brinkhaus B, Wruck K, Tag B, Mank S, Willich SN. Safety of acupuncture: results of a prospective observational study with 229,230 patients and introduction of a medical information and consent form. *Forsch Komplementmed* 2009;16(2):91-7.
91. MacPherson H, Thomas K, Walters S, Fitter M. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;323(7311):486-7.
92. White A, Hayhoe S, Hart A, Ernst E. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. *BMJ* 2001;323(7311):485-6.
93. Rao JK, Mihaliak K, Kroenke K, Bradley J, Tierney WM, Weinberger M. Use of complementary therapies for arthritis among patients of rheumatologists. *Ann Intern Med* 1999;131(6):409-16.
94. Vickers AJ, Cronin AM, Maschino AC, Lewith G, Macpherson H, Foster NE, Sherman KJ, Witt CM, Linde K. Acupuncture for Chronic Pain: Individual Patient Data Meta-analysis. *Arch Intern Med* 2012;172(19):1444-53.
95. Ratcliffe J, Thomas KJ, MacPherson H, Brazier J. A randomised controlled trial of acupuncture care for persistent low back pain: cost effectiveness analysis. *BMJ* 2006;333(7569):626.
96. Witt CM, Jena S, Selim D, Brinkhaus B, Reinhold T, Wruck K, Liecker B, Linde K, Wegscheider K, Willich SN. Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *Am J Epidemiol* 2006;164(5):487-96.
97. Yousuf S, Frick KD, Spencer C, Lao L, Berman BM, Steinwachs DM, Hochberg MC. Cost effectiveness of traditional Chinese acupuncture as adjunctive therapy in osteoarthritis of the knee. *Osteoarthritis Cartilage* 2007;15 (Suppl C):C36 (abstract #38).
98. Whitehurst DG, Bryan S, Hay EM, Thomas E, Young J, Foster NE. Cost-effectiveness of acupuncture care as an adjunct to exercise-based physical therapy for osteoarthritis of the knee. *Phys Ther* 2011;91(5):630-41.

99. Reinhold T, Witt CM, Jena S, Brinkhaus B, Willich SN. Quality of life and cost-effectiveness of acupuncture treatment in patients with osteoarthritis pain. *Eur J Health Econ* 2008;9(3):209-19.
100. Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. *BMJ* 2007;335(7616):358-9.
101. O'Connell NE, Wand BM, Goldacre B. Interpretive bias in acupuncture research?: A case study. *Eval Health Prof* 2009;32(4):393-409.
102. Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE, Stone DA, Nam BH, Kirsch I, Goldman RH. Sham device v inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006;332(7538):391-7.
103. Manheimer E, Wieland LS, Cheng K, Li SM, Shen X, Berman BM, Lao L. Acupuncture for irritable bowel syndrome. *Cochrane Database Syst Rev* 2012(5):CD005111.
104. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998;316(7127):285.

Appendix. Definitions of abbreviations and acronyms

CAM	Complementary and Alternative Medicine
CARDS	Computer Access to Research on Dietary Supplements
CBRG	Cochrane Back Review Group
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CI	Confidence Interval
CONSORT	Consolidated Standards for Reporting Trials
DO	Doctor of Osteopathic Medicine
EA	Electro-acupuncture
EULAR	The European League Against Rheumatism
FDA	US Food and Drug Administration
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
ICSI	Intracytoplasmic Sperm Injection
IPD	Individual Patient Data (meta-analysis)
IVF	In Vitro Fertilization
MCID	Minimal Clinically Important Difference
MeSH	Medical Subject Headings (in MEDLINE)
NCCAM	US National Center for Complementary and Alternative Medicine
NCI	US National Cancer Institute
NHIS	US National Health Interview Survey
NHS	UK National Health Service
NICE	UK National Institute for Health and Clinical Excellence
NIH	US National Institutes of Health
NLM	US National Library of Medicine
NR	Not Reported
NSAIDS	Nonsteroidal Anti-Inflammatory Drugs
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
OR	Odds Ratio
PLS	Plain Language Summary
RCT	Randomized Controlled Trial
RevMan	Cochrane Review Manager Software
RR	Relative Risk
SADCCT	South Asian Database of Controlled Clinical Trials
SD	Standard Deviation
SMD	Standardized Mean Difference
STRICTA	Standards for Reporting Interventions in Controlled Trials of Acupuncture
SoF	Summary of Findings (table)
TCM	Traditional Chinese Medicine
TENS	Transcutaneous Electrical Nerve Stimulation
VAS	Visual Analogue Scale
WOMAC	Western Ontario and McMaster Universities Arthritis Index

Summary

The overall purpose of this thesis has been to contribute to the systematic review evidence base for CAM interventions. As discussed in the General Introduction Chapter (Chapter 1), this is important because in this era of evidence-based medicine, physicians and policy-makers require evidence from systematic reviews to inform decisions about the value of acupuncture and other CAM interventions for treating specific health conditions. In addition, CAM interventions are widely used by the Public at considerable expense, so evidence to inform the Public's decisions about the benefits and risks of specific CAM interventions is valuable. Although the chapters that comprise this thesis all relate to the broad topic of systematic reviews and CAM, these chapters can be aggregated into two Parts. The key objective of Part 1 (Chapters 2-4) of this thesis has been to develop and analyze CAM systematic review-related databases. The key objectives of Part 2 (Chapters 5-9) have been to prepare systematic reviews to evaluate the effects of acupuncture for three specific health conditions, as well as to conduct research with a view to improve the methods of future acupuncture RCTs and systematic reviews.

The aim of Chapter 2 has been to develop a convenient way for users of *The Cochrane Library* (researchers, clinicians, consumers) to identify Cochrane reviews that are CAM-related, and to easily find Cochrane reviews on specific CAM interventions. To address this aim, a Field 'topics list' was developed, which compiled and categorized all 396 CAM-related Cochrane reviews (as of *The Cochrane Library*, Issue 4, 2009). As described in Chapter 2, the first step in developing this topics list was developing a standardized "operational" definition, to provide an objective, reproducible, and systematic method for defining and classifying CAM therapies. Chapter 2 also discusses the challenges in developing the topics list, including developing the operational definition of CAM, deciding which reviews should be included within the CAM Field's scope, developing the structured list of CAM Field-specific topics, and determining where in the topics list the reviews should be placed. Chapter 2 concludes with a discussion of remaining challenges and opportunities related to this topics list, for instance identification of gaps in Cochrane systematic review coverage of CAM therapies.

To facilitate and ease the identification of CAM trials by systematic reviewers, a specialized register of citations to CAM-related controlled trials has been developed. The first objective of Chapter 3 was to describe the sources and methods for developing the specialized register. The second objective was to analyze the trial citations included in the register, on the following characteristics: publication dates, languages, publication journals, presence in MEDINE, the type of CAM intervention addressed, and type of medical condition addressed. At the time of analysis [March 2012], the CAM Field trials register included 44,840 citations of CAM-related controlled trials. Most citations (60%) were from 2000 or later, and the majority (71%) were reported in English; the next most common language was Chinese (23%). The journals with the greatest number of citations were CAM journals published in Chinese, and non CAM nutrition journals published in English. More than one-third of the register citations (36%) were not indexed in MEDLINE. The most common CAM intervention in the register was non-vitamin, non-mineral natural products (34%), followed by Chinese herbal medicine (27%). The implications for CAM systematic reviewers of the improved accessibility of these difficult to locate CAM trials, many of which are of low methodological quality and may overestimate treatment effects, are reviewed in the Discussion section of Chapter 3.

Chapter 4 describes a review overview of all Cochrane reviews on the topic of traditional Chinese medicine (TCM). In order to identify reviews focusing on TCM, the titles and abstracts of all reviews in Issue 4, 2008 of the *Cochrane Database of Systematic Reviews* were searched. For each review, data on the number of trials included and the total number of participants were extracted. An indication of the strength of the review findings was provided by assessing the 'Authors' Conclusions' section of the systematic review abstract. This assessment of the Authors' Conclusions

was supplemented with a listing of the comparisons and outcomes showing statistically significant meta-analyses results. Seventy Cochrane reviews of TCM were identified, primarily acupuncture (n=26) and Chinese herbal medicine (n=42). The Authors' Conclusions of 7/26 acupuncture reviews and 20/42 herbal medicine reviews indicated a suggestion of benefit, which was qualified by a caveat about the poor quality and quantity of studies. The remainder of the reviews were inconclusive, due to the poor methodology and heterogeneity of the studies reviewed. The topics of the reviews that indicated a suggestion of benefit might be prioritized for future rigorous randomized controlled trials. As further discussed in the General Discussion (Chapter 10), classifying the bottom line findings of these systematic reviews was a challenging and complex endeavor that will require much additional research. Although the classification methods used in Chapter 4 may have limitations, the methods used may potentially serve to inform future research efforts to classify the overall findings of systematic reviews.

Chapter 5 is a book chapter that summarizes methodological research related to the preparation of CAM-related systematic reviews, and also suggests approaches for addressing some of the challenges in systematically reviewing CAM interventions. Chapter 5 begins by explaining how systematic reviews ensure rigorous quality standards and maintain objectivity during each phase of review preparation, including: (1) identifying relevant RCTs; (2) assessing the quality of the RCTs; and (3) combining the data from the RCTs. The discussion of assessing quality of RCTs (section 2) is related to thesis Chapter 6, and section 2 discusses the advantages and limitations of contacting RCT authors to request information about randomization and blinding procedures not described in RCT publications. The discussion of combining data from RCTs (section 3) uses the acupuncture for low back pain review (Chapter 7) as a case study to discuss the decisions that systematic reviewers need to make in terms of whether and how to statistically pool the results of similar but separate RCTs. The remainder of Chapter 5 focuses on the advantages and drawbacks of different methods for assessing treatment adequacy and practitioner adequacy in CAM trials. Systematic review methodology is evolving, and advice for addressing methodological challenges can change in light of new research findings, or in light of additional experience gained in preparing systematic reviews. For example, in Chapter 5, a holistic approach to assessing treatment adequacy had been advocated. However, as described in the General Discussion (Chapter 10), the experience in preparing the three systematic reviews comprising this thesis led to a revision of this earlier view about the optimal approach for assessing treatment adequacy. Namely, the additional experience gained has led to the conclusion that an objective method of assessing individual treatment adequacy-related components has advantages over a holistic, composite approach of assessing treatment adequacy.

Chapter 6 describes a survey undertaken to evaluate whether the description of randomization and blinding in published articles is an accurate and complete reflection of the study procedures used. This question is important to address because systematic reviews generally evaluate RCTs based on the published articles. A major disadvantage of quality assessment based solely on the published articles is that this relies on a critical assumption that what is written in the article reflects actual procedure and that "if it is not reported, it probably was not done". For this survey, the authors of 51 RCTs included in a systematic review of acupuncture for chronic pain were sent a brief questionnaire that included questions related to the following three study quality dimensions: (1) generation of allocation sequence, (2) allocation concealment, and (3) blinding of outcomes assessor. Among the 51 questionnaires sent out, 35 responses were received. Of 35 studies described as randomized in the published articles, associated survey responses indicated that four actually used quasi-randomized methods (i.e., methods of assignment not considered strictly random, such as assignment by alternation, date of birth, or medical record number). Among published articles with missing information on allocation concealment and outcomes assessor blinding, 13 of 34

studies used adequate allocation concealment and two of 10 studies were blinded, according to survey responses. With the exception of the four quasi-randomized trials described as randomized in the RCT publications, survey responses generally confirmed information about randomization and blinding already described in the investigators' RCT publications. Thus, this survey suggested that contacting trial authors may result in obtaining previously unpublished information about methodological quality. However, the potential gains of obtaining the missing information need to be weighed against the reporting bias such efforts may introduce, particularly when asking investigators to select from a list of response choices, as was done in this survey.

Chapter 7 describes a systematic review with the objective of assessing acupuncture's effectiveness for treating low back pain. For this systematic review, computerized databases were searched to find RCTs that assigned patients with low back pain to receive either acupuncture or one of the following comparison treatments: sham acupuncture, sham transcutaneous electrical nerve stimulation (TENS), no low back pain treatment, or traditional treatments for low back pain. The 33 RCTs that met inclusion criteria were subgrouped according to acute or chronic pain, style of acupuncture, and type of control group used. The primary outcome of this review was a quantitative synthesis of the short-term effectiveness of acupuncture on pain in each subgroup. For the quantitative synthesis, the RCTs' standardized mean differences (SMDs) were used. These SMDs were calculated as differences in post-treatment scores between groups. For the primary outcome of short-term relief of chronic pain, the meta-analyses showed that traditional Chinese acupuncture was significantly more effective than sham acupuncture, sham TENS, and no additional treatment. For patients with acute low back pain, data were sparse and inconclusive. Data were also insufficient for drawing conclusions about acupuncture's short-term effectiveness compared with most other therapies. The overall conclusion that acupuncture may be more effective than a sham acupuncture control for providing short-term relief of chronic low back pain was considered preliminary. This was because the comparison of acupuncture versus sham acupuncture was based on only four small to moderately sized trials, and also because there were several large and recently completed but not yet published trials that would strongly influence these findings

Chapter 8 describes a Cochrane review with the objective of assessing acupuncture's effectiveness for treating peripheral joint osteoarthritis. For this review, computerized databases were searched to find RCTs that compared needle acupuncture with a sham, another active treatment, or a waiting list control group in people with osteoarthritis of the knee, hip, or hand. Authors of all included RCTs were contacted to obtain additional information about their RCTs' methods and results. For the primary outcomes of short-term improvements in pain and functioning, the effects of acupuncture relative to control were calculated as SMDs. These SMDs were calculated using the differences in improvement between groups. Sixteen RCTs involving 3,498 people were included in the systematic review. Twelve of the RCTs included only people with osteoarthritis of the knee, three only osteoarthritis of the hip, and one a mix of people with osteoarthritis of the hip and/or knee. In comparison with a sham acupuncture control, true acupuncture resulted in statistically significant, short-term improvements in osteoarthritis pain and function. However, these pooled short-term improvements did not meet our predefined thresholds for clinically relevant improvements, and furthermore, the trials that were pooled showed substantial statistical heterogeneity. This statistical heterogeneity among the pooled trials was largely explained by the adequacy of participant blinding variable, which showed a statistically significant subgroup effect. When restricting to only those sham-controlled trials that were judged to have adequate participant blinding, the pooled improvements were no longer statistically significant. Trials with a waiting list control showed statistically significant and clinically relevant benefits. Based on these findings, it was concluded that the use of different control comparators

explains the variability in the results of acupuncture for osteoarthritis trials, and that placebo effects due to patient expectations probably accounts for most of the observed benefits.

Chapter 9 describes a systematic review with the objectives of assessing adjuvant acupuncture's overall pooled effect in increasing in vitro fertilization (IVF) pregnancy success rates, and also assessing whether study design-, treatment-, and population-related factors influence these effect estimates. For this review, computerized databases were searched to find RCTs that compared needle acupuncture administered within one day of embryo transfer, versus sham acupuncture or no adjuvant treatment. The investigators of all RCTs provided additional methodological details and outcome data not included in their original RCT publications. The primary outcome was clinical pregnancy rates. Eleven subgroup variables (five clinical and six methodological) were pre-specified to investigate sources of heterogeneity, using single covariate meta-regressions. Sixteen trials (4,038 participants) were included in the meta-analyses. There was no statistically significant difference between acupuncture and control when combining all trials, or when restricting to sham-controlled or no adjuvant treatment-controlled trials. The type of control used did not significantly explain the statistical heterogeneity. However, the no adjuvant treatment-controlled trials had a somewhat larger pooled effect than the sham-controlled trials, and the pooled effect of the no adjuvant treatment-controlled trials was statistically significant if an outlying trial was removed. Of the eleven variables pre-specified for meta-regression subgroup analyses, only the baseline (i.e., control group) pregnancy rate variable showed a statistically significant subgroup effect, and this variable explained most of the heterogeneity in results for the primary outcome measure across all trials. The subgroup finding of a benefit in trials with lower, but not higher, baseline pregnancy rates was the only statistically significant subgroup finding in our earlier review, was confirmed in this update, and was not explained by any confounding variables evaluated. However, we concluded that this subgroup benefit finding still requires a cautious interpretation because of the multiple subgroup tests conducted, the risk of unidentified confounders, the multiple different factors that determine baseline rates, and the possibility of publication bias.

This thesis concludes with a General Discussion (Chapter 10), which reflects on the findings and discusses the methodological issues that were identified while working on this thesis, and sets the agenda for future research after the completion of this thesis. Part 1 of Chapter 10 discusses the challenges of developing eligibility criteria for including studies in the databases of CAM-related reviews and trials described in Chapters 2-4. Part 1 also discusses the challenges of developing criteria for further classifying the trials and reviews selected for inclusion in these databases. Part 2 discusses Chapters 5-9, the systematic reviews that were conducted to evaluate the effects of acupuncture for three specific conditions, and methodological studies conducted in the context of these evaluations. A primary objective in preparing these three reviews (or indeed any systematic reviews), was to determine whether acupuncture 'works' for treating the given conditions. This determination was made based not only on the meta-analysis estimates from the pooled trials, but also the various biases of the trials, and the possibility of different effects across different subgroups of trials. Part 2 discusses the primary issues that were encountered in attempting to draw conclusions about the effects of acupuncture relative to a given control for the three conditions studied, as well as methodological problems of existing trials which impede the ability to draw conclusions about whether acupuncture 'works' relative to a given control. Some issues discussed in Part 2 include the calculation, interpretation, and reporting of subgroup analyses (particularly as this relates to the use of sham acupuncture versus usual care as the control intervention) and sensitivity analyses, and the methodology of the pre-requisite step for these subgroup and sensitivity analyses, which is contacting the trial authors to obtain the necessary data. Other issues discussed in Part 2 of Chapter 10 include the appraisal of the effects of the unpublished or not yet published, trials, and the interpretation of the clinical relevance of the treatment effects for real

world practice. Part 2 concludes with a brief, general overview of implications for future research of systematic reviews of acupuncture. This section largely focuses on the advantages and drawbacks of sham-controlled efficacy trials versus comparative effectiveness pragmatic trials in investigating the effects of acupuncture.

Samenvatting

Systematische reviews op het gebied van complementaire en alternatieve geneeskunde: belang, methoden en voorbeelden uit de acupunctuur

Dit proefschrift heeft ten doel bij te dragen aan het inzicht in de effectiviteit van complementaire en alternatieve geneeskunde (CAG) op basis van systematische reviews. Zoals besproken in de algemene inleiding (Hoofdstuk 1), is dat van belang omdat in dit tijdperk van 'evidence based medicine' artsen en beleidsmakers systematische reviews willen gebruiken om de waarde van acupunctuur en andere CAG interventies bij specifieke gezondheidsproblemen te kunnen inschatten. CAG interventies worden vaak toegepast en brengen substantiële kosten met zich mee. Daarom is het van belang dat er een goede afweging van de effecten en de risico's kan worden gemaakt. Alle hoofdstukken van dit proefschrift hebben betrekking op het brede onderwerp CAG en systematische reviews, maar het proefschrift bestaat uit twee delen. Deel 1 (Hoofdstuk 2 t/m 4) richt zich op de vraag hoe databases betreffende systematische reviews naar CAG het beste kunnen worden ontwikkeld en geanalyseerd. Deel 2 (Hoofdstuk 5 t/m 9) richt zich op systematische reviews van acupunctuur bij drie specifieke gezondheidsproblemen, alsmede op onderzoek gericht op verbeteringen in de methodologie van gerandomiseerde klinische studies en systematische reviews op het terrein van acupunctuur.

Hoofdstuk 2 heeft als doel een handige methode te ontwikkelen waarmee gebruikers van de Cochrane Library (onderzoekers, clinici en patiënten) systematische reviews over CAG kunnen identificeren, ook voor specifieke gezondheidsproblemen. Hiertoe werd een lijst van onderwerpen opgesteld waarmee vervolgens alle 396 CAG-gerelateerde Cochrane reviews (Cochrane Library Issue 4, 2009) werden gecategoriseerd. De eerste stap bij het opstellen van de lijst van onderwerpen was het ontwikkelen van een operationele definitie van CAG, zodat CAG interventies objectief, reproduceerbaar en systematisch kunnen worden geclassificeerd. Hoofdstuk 2 beschrijft enkele uitdagingen bij het ontwerpen van deze lijst van onderwerpen, inclusief het formuleren van de operationele definitie, het besluiten welke reviews wel en welke niet tot het CAG veld behoren, het opstellen van de lijst van onderwerpen en het bepalen welke reviews bij welk onderwerp van de lijst passen. Het hoofdstuk sluit af met een discussie over de wijze waarop deze lijst van onderwerpen verder kan worden gebruikt, bijvoorbeeld om te laten zien over welke CAG interventies er nog geen Cochrane review beschikbaar is.

Om auteurs van systematische reviews van CAG interventies te helpen, werd een gespecialiseerd register van citaties naar CAG-gerelateerde gerandomiseerde klinische studies (trials) opgezet. Het eerste doel van Hoofdstuk 3 was om de bronnen en methoden te beschrijven waarmee dit register werd ontwikkeld. Het tweede doel was om de trial citaties te analyseren aan de hand van de volgende kenmerken: publicatiedatum, taal, tijdschrift van publiceren, aanwezigheid in MEDLINE, type CAG interventie, en type gezondheidsprobleem. Ten tijde van de analyse (maart 2012) bevatte het register 44.840 citaties naar CAG-gerelateerde trials. De meeste citaties (60%) waren van 2000 of later en de meeste citaties waren in het Engels (71%), met Chinees als de daarop volgende taal (23%). De meeste citaties betroffen Chinese tijdschriften gericht op CAG, gevolgd door Engelse tijdschriften gericht op voeding. Meer dan een derde (36%) van de citaties uit het register waren niet geïndexeerd in MEDLINE. De meest voorkomende interventie (34%) betrof natuurlijke producten (uitgezonderd vitamines en mineralen), gevolgd door Chinese kruidengeneeskunde (27%). Het register maakt een verbeterde toegang tot CAG trials mogelijk en in de discussion van dit hoofdstuk wordt ingegaan op de implicaties hiervan, inclusief het feit dat

veel trials van geringe methodologische kwaliteit zijn en het effect van de betreffende interventie in de regel zullen overschatten.

Hoofdstuk 4 geeft een overzicht van alle Cochrane reviews over traditionele Chinese geneeskunde. Daartoe werden alle titels en samenvattingen van de Cochrane Library (Issue 4, 2008) doorzocht. Van elke review werd het aantal trials en het totaal aantal deelnemers genoteerd. Een indicatie van de sterkte van het bewijs werd verkregen door de conclusies in de samenvatting van de reviews te beoordelen. Daarnaast werd een lijst opgesteld van vergelijkingen tussen interventies en de uitkomsten die een statistisch significant resultaat lieten zien in de meta-analyse. Er werden 70 Cochrane reviews van traditionele Chinese geneeskunde geïdentificeerd, die met name acupunctuur (n=26) of Chinese kruidengeneeskunde (n=42) betroffen. De conclusies uit de samenvattingen duidden op een positief effect in 7/26 acupunctuur reviews en in 20/42 kruidengeneeskunde reviews, maar zowel de kwaliteit als de omvang van de trials beschreven in deze reviews was gering. Uit de overige reviews konden door de slechte methodologie en de grote heterogeniteit geen conclusies worden getrokken. De onderwerpen van de reviews waarin een positief effect werd gesuggereerd verdienen prioriteit in toekomstig onderzoek. De classificatie van de conclusies uit de reviews van traditionele Chinese geneeskunde was een complexe onderneming waarnaar nog veel nader onderzoek nodig is. In hoofdstuk 10 wordt hierop verder ingegaan. Ondanks de beperkingen van de in hoofdstuk 4 toegepaste classificatiemethode, lijkt deze in potentie bruikbaar bij toekomstige inspanningen om de conclusies van systematische reviews te classificeren.

Hoofdstuk 5 is een hoofdstuk uit een boek dat een samenvatting geeft van methodologisch onderzoek gericht op de voorbereiding van systematische reviews op het terrein van CAG. Het bevat ook suggesties hoe om te gaan met een aantal uitdagingen die zich voordoen bij het maken van dergelijke systematische reviews. Het hoofdstuk begint met een uitleg hoe systematische reviews zorgen voor een hoge kwaliteitsstandaard en voor objectiviteit gedurende alle fases van het uitvoeren van een systematische review: (1) identificatie van de relevante trials, (2) beoordelen van de kwaliteit van de trials, en (3) het combineren van de resultaten over de effectiviteit van interventies gerapporteerd in trials. De discussie over de beoordeling van de kwaliteit van trials gaat met name in op de voor- en nadelen van het opvragen van nadere informatie over randomisatie en blindering bij de auteurs van trial publicaties. De discussie over het combineren van gegevens betreffende effectiviteit van interventies gaat met name in op de besluitvorming rond het al dan niet statisch poolen van de gegevens en de wijze waarop dit eventueel kan worden gedaan. De rest van hoofdstuk 5 gaat in op de voor- en nadelen van verschillende methoden om de kwaliteit van de behandeling en van de therapeuten in trials te beoordelen. De methodologie van systematische reviews is nog volop in ontwikkeling. Daardoor kunnen adviezen over hoe om te gaan met methodologische uitdagingen veranderen in het licht van nieuwe ervaringen en resultaten van onderzoek. In hoofdstuk 5 wordt bijvoorbeeld uitgegaan van een holistische benadering om de kwaliteit van de behandeling te beoordelen. Maar in de algemene discussie (Hoofdstuk 10) blijkt dat de ervaring met het uitvoeren van de drie systematische CAG reviews heeft geleid tot een herziening van dit advies. De conclusie in hoofdstuk 10 is dat een meer objectieve methode gericht op de verschillende kwaliteitscomponenten van een behandeling de voorkeur verdient.

Hoofdstuk 6 beschrijft een onderzoek naar de vraag of de beschrijving van de randomisatie en de blindering in gepubliceerde artikelen een goede en complete weergave geeft van de gehanteerde procedures. Deze vraag is van belang omdat in de regel in een systematische review wordt afgegaan op hetgeen er in gepubliceerde artikelen over de trials wordt gezegd. De onderliggende

aanname is dat dit de in werkelijkheid gevolgde procedures goed weergeeft en dat als er iets niet wordt gerapporteerd, het vermoedelijk ook niet is gedaan. Voor dit onderzoek werden de auteurs benaderd van 51 trials die waren opgenomen in een systematische review van acupunctuur voor chronische pijn. Zij ontvingen een korte vragenlijst die inging op drie dimensies van validiteit van onderzoek: (1) het genereren van de toewijzingsvolgorde, (2) het blinderen van deze toewijzingsvolgorde, en (3) de blindering van degene die de uitkomst vaststelde. Er werd een reactie ontvangen van de auteurs van 35 van de 51 trials. Van deze 35 trials, die in de publicaties allemaal als gerandomiseerd werden beschreven, bleken er 4 van quasi-randomisatie methoden gebruik te hebben gemaakt. Van de 34 trials met ontbrekende informatie over het blinderen van de toewijzingsvolgorde (2) bleken er 13 toch een adequate methode te hebben gehanteerd. En van de 10 trials met ontbrekende informatie over degene die de uitkomst vaststelde (3), bleken er 2 toch wel geblindeerd te zijn. Verder werden de onderzoeksprocedures zoals beschreven in de publicaties over het algemeen bevestigd. De resultaten van dit onderzoek suggereren dat het de moeite waard kan zijn om bij ontbrekende informatie contact op te nemen met de auteurs van de trials. Maar de potentiële voordelen moeten daarbij worden afgewogen tegen de mogelijke nadelen, met name de kans op vertekening doordat auteurs wellicht gewenste antwoorden geven. Dit kan met name een rol spelen wanneer gekozen moet worden uit een lijst mogelijke opties, zoals werd gedaan in het hier beschreven onderzoek.

Hoofdstuk 7 beschrijft een systematische review naar de effectiviteit van acupunctuur bij de behandeling van lage rugpijn. Hiervoor werden meerdere digitale bibliografische bestanden doorzocht om trials te vinden waarin patiënten waren toegewezen aan een acupunctuurbehandeling of aan een van de volgende andere interventies: placebo acupunctuur, placebo transcutane elektro neurostimulatie (TENS), geen behandeling of de gebruikelijke zorg voor lage rugpijn. De 33 trials die voldeden aan de insluitcriteria werden onderverdeeld naar acute of chronische pijn, type acupunctuur en type vergelijkingsgroep. De primaire uitkomst van deze systematische review was een kwantitatieve synthese van de korte termijneffecten van acupunctuur op pijn. Bij patienten met chronische lage rugpijn blijkt Chinese acupunctuur significant effectiever dan placebo acupunctuur, placebo TENS, en geen behandeling. Voor acute lage rugpijn waren de gegevens schaars en konden geen conclusies worden getrokken. Hetzelfde gold voor de korte termijn vergelijking met de meeste andere behandelingen. Er wordt gewaarschuwd dat de conclusie dat acupunctuur een korte termijneffect heeft bij chronische lage rugpijn voorbarig kan zijn, met name omdat de vergelijking met placebo acupunctuur slechts 4 kleine tot middelgrote trials betrof. Ten tijde van het uitvoeren van de review waren een aantal grote trials recent afgesloten doch nog niet gepubliceerd, en de resultaten van deze trials zouden een grote impact kunnen hebben op de conclusie van deze systematische review.

Hoofdstuk 8 beschrijft een systematische review naar de effectiviteit van acupunctuur bij artrose van perifere gewrichten. Hiervoor werden digitale bibliografische bestanden doorzocht gericht op het vinden van trials waarin acupunctuur werd vergeleken met placebo, met een andere actieve behandeling, of met een wachtlijstcontrole bij patiënten met artrose van de knie, heup of hand. Er werd met de auteurs van ingesloten trials contact opgenomen om aanvullende informatie over methoden en resultaten te verkrijgen. De primaire uitkomst betrof het verschil in korte termijn verbeteringen van pijn en functioneren. Zestien trials met in totaal 3.498 deelnemers werden betrokken in deze systematische review. Twaalf trials betroffen patiënten met artrose van de knie, 3 artrose van de heup, en 1 betrof een mix van heup en knie artrose. In vergelijking tot

placebobehandeling werd een statistisch significant voordeel gevonden voor acupunctuur wat betreft pijn en functioneren. Dit verschil was echter klein, en overschreedt niet de minimale waarde voor klinische relevantie. Er was tevens sprake van substantiële statistische heterogeniteit, dat wil zeggen grote verschillen tussen de trials wat betreft de schatting van het effect van acupunctuur. In een subgroep analyse werd gekeken naar het de uitkomsten in trials waarin de blindering van deelnemers naar alle waarschijnlijkheid succesvol was verlopen. In deze subgroep analyse bleek het korte termijn effect van acupunctuur kleiner en niet significant. Trials met een wachtlijstcontrole lieten een statistisch significant en klinisch relevant effect zien in het voordeel van acupunctuur. Op basis van deze bevindingen werd de conclusie getrokken dat het gebruik van verschillende vergelijkinggroepen de inconsistentie in de uitkomsten van trials naar het effect van acupunctuur bij artrose kan verklaren. Het feit dat in trials waarin blindering van deelnemers niet succesvol of niet mogelijk is (bijvoorbeeld bij gebruik van een wachtlijstcontrole) grotere effecten worden gevonden geeft aan dat verwachtingen van patiënten een belangrijke determinant zijn van de waargenomen effecten van acupunctuur.

Hoofdstuk 9 beschrijft een systematische review naar de effecten van acupunctuur als aanvullende behandeling bij in-vitro fertilisatie, waarbij de kans op zwangerschap de belangrijkste uitkomstmaat was. Er werd ook gekeken naar de invloed van verschillende variabelen op de uitkomsten, waaronder aspecten van de opzet van de trial, de behandeling en de onderzoekspopulatie. Voor deze systematische review werden digitale bibliografische bestanden doorzocht op trials die acupunctuur binnen één dag na de embryo transfer vergeleken met placebo acupunctuur of geen aanvullende behandeling. De auteurs van alle trials verschaften aanvullende informatie over de methoden en uitkomsten van hun trial die niet waren beschreven in de oorspronkelijke publicaties. Voor 11 vooraf gekozen variabelen (5 klinische en 6 methodologische) werd de invloed op de uitkomsten bestudeerd door middel van metaregressie analyse. Zestien trials (4.038 deelnemers) werden betrokken in de systematische review. Er was geen statistisch significant verschil tussen acupunctuur en de vergelijkingsgroep wanneer alle trials werden gecombineerd of wanneer afzonderlijk naar de vergelijking met placebo acupunctuur of geen aanvullende behandeling werd gekeken. Het type vergelijkingsgroep kon de statistische heterogeniteit niet verklaren, hoewel wel een iets groter gepooled effect werd gevonden voor trials waarin acupunctuur werd vergeleken met geen aanvullende behandeling, en dit effect bleek statistisch significant wanneer één afwijkende studie werd weggelaten. Van de 11 variabelen die vooraf voor metaregressie analyse waren geselecteerd, liet alleen de zwangerschapkans in de vergelijkingsgroep een statistisch significant effect zien. Deze variabele verklaarde tevens de meeste variatie in de primaire uitkomst. De bevinding dat acupunctuur een gunstig effect heeft met name in populaties met een lagere zwangerschapskans werd ook in een eerdere systematische review gevonden en kon niet door andere variabelen worden verklaard. Er wordt echter gepleit voor een terughoudende interpretatie van deze uitkomst vanwege het feit dat veel subgroep analyses werden uitgevoerd (verhoogde kans op een toevalsbevinding) en dat andere, niet gemeten variabelen het effect zouden kunnen verklaren.

Het proefschrift wordt afgesloten met een algemene discussie (Hoofdstuk 10), waarin op de bevindingen wordt gereflecteerd, verschillende methodologische kwesties nader worden besproken, en de agenda voor toekomstig onderzoek wordt geformuleerd. Deel 1 van dit hoofdstuk bespreekt de uitdagingen van het ontwikkelen van insluitcriteria voor digitale databases van trials en systematische reviews op het terrein van CAG, zoals beschreven in deel 1 van de dissertatie (Hoofdstukken 2 t/m 4). Tevens wordt besproken hoe deze trials en systematische reviews het beste

kunnen worden geclassificeerd. Deel 2 richt zich op de hoofdstukken 5 t/m 9, de systematische reviews die werden verricht om het effect van acupunctuur bij drie specifieke gezondheidsproblemen te schatten met de bijbehorende methodologische studies. Het primaire doel van deze reviews, zoals van elke systematische review, was om vast te stellen of acupunctuur 'werkt' bij deze aandoeningen. Daarbij werd niet alleen naar de schattingen uit de gepoolde analyses gekeken, maar tevens naar mogelijke bronnen van vertekening en eventuele verschillen tussen subgroepen. Deel 2 van Hoofdstuk 10 bespreekt de kwesties die zich voordeden bij het trekken van conclusies over de effecten van acupunctuur ten opzichte van een vergelijkingsgroep. Tevens komen de methodologische beperkingen van de trials aan de orde en de gevolgen daarvan voor het trekken van conclusies over het effect van acupunctuur. In dit verband komen onder meer de analyse, interpretatie en rapportage van subgroep analyses aan de orde, in het bijzonder het gebruik van placebo acupunctuur versus gebruikelijke zorg als vergelijkingsgroep, alsmede de methodologie en het vooraf kiezen van variabelen voor subgroep- en sensitiviteitsanalyses. Ook wordt in deel 2 van Hoofdstuk 10 besproken hoe om te gaan met ongepubliceerde en nog niet gepubliceerde trials, en met de interpretatie van uitkomsten in termen van klinische relevantie met oog op de toepasbaarheid van resultaten voor de klinische praktijk. Het hoofdstuk wordt afgesloten met een overzicht van de implicaties van de bevindingen voor toekomstig onderzoek naar systematische reviews van acupunctuur. Daarbij wordt met name ingegaan op de voor- en nadelen van placebo vergelijkingsgroepen versus vergelijkingsgroepen met een actieve behandeling.

Acknowledgments

Lex Bouter: I will always be grateful for the opportunity you have provided me in my attempt to obtain a PhD, as well as for your profoundly wise advice on the component chapters of the PhD thesis. Putting together the final thesis book recently has reminded me again of all that has been accomplished under your supervision, starting in 2006, with our work on the 2 reviews that served as preliminary work for thesis chapters 8 and 9. As you may recall, I had difficulty addressing some issues during the final stages of the publication process for these 2 reviews. Despite all your responsibilities, you had responded to my time-sensitive e-mails with sound advice, often on the weekends or evenings. I believe that following your advice was instrumental in the final acceptance of these 2 early reviews, which also subsequently led to my official acceptance into the EMGO PhD program. Your guidance continued with the manuscripts of the thesis chapters, where you typically provided the first comments and revisions on the versions of the manuscripts e-mailed to co-authors. Your initial comments typically set the stage for the subsequent revisions of the manuscripts, with Danielle and other authors substantially adding to and building upon your comments or even suggesting some alternatives. As a result of contemplating and studying all of your comments, remarks, and revisions, I believe that I have improved my writing, such that it is now clearer, more concise, and more focused. As one example, the Introduction and Discussion Chapters of the thesis were particularly challenging for me, in synthesizing a large number of complex ideas in a concise way. However, these were the chapters in which I probably learned the most about how to separate the key points from the extraneous points, and how to write about these key points in a more concise and focused manner. Indeed, when I first received your e-mail indicating that my first draft of the General Discussion Chapter needed to be shortened by about 50%, I was uncertain how I would accomplish that. However, after careful consideration and work on the subsequent revision, I realized that the sections that you and Danielle pointed out were not critically important could, and should, be removed. If this General Discussion Chapter had not been greatly shortened, any reader would likely have had a difficult time getting through it. This is just one example among many of how the guidance received from you and Danielle has improved my capability to separate the 'wheat from the chaff', in succinctly writing up a paper. Learning to give more thought to the relevance of potential ideas that could be introduced, during the writing process itself, rather than at its completion, has been a particularly valuable lesson to learn, as this is a much more efficient approach than spending the time to write up text that will then be cut at the end. Finally, I am indebted to you for your generosity of your valuable time in reviewing and commenting on the thesis chapters. I have tried to do my best to reciprocate by getting the subsequent revisions of the chapters in the best shape that I could before sending them back to you and Danielle, and also by doing my best to comprehensively address all of your previous comments, so that your time would be spent most efficiently.

Danielle van der Windt: I am forever grateful to you for providing me with this opportunity. I have benefited immensely from your hands-on guidance, mentorship, and insight, as well as the many valuable lessons you have taught me in researching, writing, and publishing one's work. Both you and Lex have been so generous with your time and guidance, and I only hope I can at least somewhat reciprocate your generosity at some point in the future. As you will recall, your involvement began in late 2007, with your helpful guidance in responding to the Editors'/peer-reviewers' comments on the February 2008 acupuncture and IVF review (preliminary work for Chapter 9). At that time, you had indicated that you did not need to be added as an author on that 2008 IVF review, in spite of your contribution during the peer-review process, because you were not an author at the time the paper was submitted. I was delighted therefore that you were involved, from the outset, with the subsequent 2010 Cochrane review of acupuncture for osteoarthritis (Chapter 8). Among your many contributions to this review, perhaps the most valuable was your

resolving differing opinions among co-authors about the overall Conclusions. Through your diplomacy, tact, and optimal wording, all co-authors were able to agree on the Conclusions you suggested, despite some previous differences among co-authors on the optimal wording. For the subsequent bibliometric analysis (Chapter 3), the entire Discussion section and the analysis and interpretation of the data included therein, as well as parts of the other sections, were written and rewritten largely based on the extensive and valuable comments and remarks of you and Lex. This resulted in a substantially improved paper. For the 2013 adjuvant acupuncture for IVF review (Chapter 9), your suggestions on how to organize the Discussion section, present the Results tables, and conduct the meta-regressions were invaluable, and greatly improved the paper. During the peer-review process for this paper, your suggestions and revisions urging that we provide clear and focused arguments for aspects that we did not want to change, while making revisions based on reasonable changes, was clearly the best way forward. This was a valuable lesson to learn, and will be the approach I will take for future peer-review responses. I believe that your revisions and comments on the 2013 IVF manuscript, as well as your contributions to the peer-review response (together with Jayne and Pat), may well have made the difference between acceptance and rejection of this paper. Finally, I was especially grateful for your patience in reviewing the multiple drafts of the General Introduction and General Discussion Chapters, and for yours' and Lex's extensive comments and revisions, which I believe resulted in each subsequent revision representing an improvement from the previous version.

Patricia Langenberg: Your mentorship, guidance, and support over the years have been invaluable, starting with the classes you taught in my Master's program at the University of Maryland, Baltimore, and continuing with your guidance on thesis chapters 7 and 9. For thesis Chapter 7, as you may recall, the outcome data of these acupuncture for low back pain trials were analyzed for the publication using post-treatment values. A subsequent Letter to the Editor suggested reanalyzing the trials using change from baseline values, to see if the results of the true versus sham acupuncture comparison would be consistent. I am grateful for your having guided me in responding to this Letter to the Editor by overseeing the re-analysis of the sham acupuncture-controlled trials using the change from baseline values. This involved some advanced statistics to calculate the best estimates of between group change scores from these incompletely reported trials (as also discussed in the thesis General Discussion Chapter), and without your advice and guidance on these complex issues, the Letter to the Editor would have been much more difficult to address. Your invaluable guidance on complex methodological issues continued with the 2013 IVF review (Chapter 9), with your oversight of the meta-regression analyses and your advice on interpreting the small study effects observed. Finally, I will always be grateful to both you and Lex for your wisdom and good judgment in helping to guide me through the review and publication of the 2008 acupuncture and IVF review (i.e., preliminary work for thesis Chapter 9), particularly at the final stages of review.

Iain Chalmers: Thank you so much for taking the time to thoroughly review my thesis manuscript, and for providing helpful suggestions for improvement and clarification. Your time and contributions in reviewing the work are particularly appreciated given your skepticism about the interventions. As a result of your suggestions, important revisions and clarifications were made to 5 of the thesis chapters (i.e., Introduction, Discussion, Summary, Chapter 3, Chapter 9), before the thesis book was bound for eternity. Your suggestions were received in the nick of time to also make changes to the journal article versions of thesis chapters 3 and 9. For Chapter 9 in particular, your suggestions resulted in important changes, which improved the clarity and comprehensibility of the publication. Your comments and suggestions not only improved and clarified aspects of the thesis

manuscripts, but also researching and addressing the key points you raised helped me get a head-start in preparing for the viva.

Brian Berman: Thank you for hiring me as the Cochrane Complementary Field coordinator/methodologist in January 2003, and also for your pioneering work in CAM. I am particularly grateful to you for first opening up this possibility of my obtaining a PhD, by introducing me to Lex at the 2005 Cochrane Colloquium. I am also thankful for your having encouraged me not to abandon the PhD idea, during a challenging time in 2009 when I was considering doing so. I am happy that I stuck with it, and appreciate your having encouraged me to do so.

Kay Dickersin: Thank you for hiring me as the US Cochrane Center coordinator/methodologist in 1997, for introducing me to the importance of systematic reviews, trial registers, and evidence-based healthcare, and for your pioneering work and tenacity in advancing the ideas of publication bias and trial registration.

Lixing Lao: Thank you for having rated the adequacy of the acupuncture treatments and the credibility of the sham controls and their risks of affecting the outcomes, while blinded to the identities and results of the studies, for thesis Chapters 8 and 9. Your assessments provided a methodological addition which was valuable in the interpretation of both of these reviews. For example, in the acupuncture for osteoarthritis Cochrane review (Chapter 8), your assessments of both the credibility and likely physiological activity of the sham acupuncture control interventions resulted in the finding that the six trials that were judged to have been successfully blinded were the same six trials judged to have used a sham control with potential physiological activity which might have affected the outcomes, which was a highly important finding in the overall interpretation of the review. You have also been a greatly valued collaborator and colleague for the past 11 years. Congratulations on your recent appointment as the professor and director of the school of Chinese medicine at the University of Hong Kong. I hope to be able to visit you there in the future.

Jeanette Ezzo: I am exceedingly grateful to have had the opportunity to collaborate with you on thesis chapters 5 and 6, as well as for our other past and ongoing collaborations on Cochrane reviews. Because you are so careful, accurate, thoughtful, and balanced, I consider your systematic reviews to be the 'gold standard', and the model that I strive to emulate. With all of your roles and responsibilities, including your work at the mission and your massage practice, your work in preparing Cochrane reviews in your free time really demonstrates your commitment to the Cochrane Collaboration and your commitment to also improving peoples' well-being on a population level through the preparation of Cochrane reviews to best inform clinical practice.

Susan Wieland: You have been such a valued collaborator and colleague over the past ~15 years. I am so pleased with the work we have done together in furthering the infrastructure of the Cochrane CAM Field, by our development of the CAM Field topics list and specialized register of CAM trials, as described in thesis Chapters 2 and 3, respectively. It would be so great if you were able to make it to Amsterdam for the viva.

Ke Cheng: Thank you for having provided great support with data checks, running analyses, and helping with the figure/table preparation, for thesis Chapters 4, 8, and 9, while you were a visiting scholar at the CAM Field. I look forward to our continued collaborations, and to seeing through to

completion your Cochrane reviews of acupuncture for hip osteoarthritis and acupuncture for acute low back pain.

Ellis Verkerk: I would like to thank you for making it possible for me to complete a PhD in a foreign country. I had so many questions, and you knew the answers (or found out the answers) to all of them. Your generosity and kindness in cheerfully offering your support in helping to coordinate the planning of the viva and the printing and distribution of the thesis books have been very much appreciated.

Ellen Visser: Thank you for coordinating the early administrative activities involved in my PhD program. Your steadfast help with the formalities of applying for formal admittance to the EMGO PhD program has been very much appreciated.

I would like to thank my parents, Wallace and Eileen Manheimer, for their keen enthusiasm to go to Amsterdam to be at the viva, and also for inviting me to visit them for short holidays/long weekends, which were welcome breaks from the thesis work. I would also like to thank my sister, Karen Hart, my brother-in-law Neil Hart, and their 3 wonderful children, Clara-Rose, Zachary, and Isaac for planning to attend the viva. I am so delighted that you are all willing and able to travel to Amsterdam and provide your support at this major event in my life. I would especially like to thank Eric Vernon for all your support, and for having stuck it out with me during this busy time of trying to earn a PhD. Despite our busy schedules, we still always managed to have lots of great times during our free time.

I would like to thank the following co-authors of the trials included in the thesis systematic reviews, for confirming and providing data related to their respective trials: Mary Arnoldi and Luciana de Lauretis (for the Arnoldi et al. trial), Monica Benson (for the Fratterelli et al. trial), Daniela Braga (for the Madaschi et al. trial), Birgitte Christensen, LaTasha Craig, Stefan Dieterle, Alice Domar, Luca Fallo (for the Omodei et al. trial), Elisabetta Feliciani, Nadine Foster, Elisabeth C. Larsen (for the Andersen et al. trial), Irene Moy, Ernest Ng (for the 2009 and 2010 So et al. trials), Wolfgang Paulus (for the 2002 and 2003 Paulus et al. trials), Caroline Smith, Jorge Vas, Lars Westergaard, Lynn Williamson, Claudia Witt, and Steffen Witte.

I would like to thank the following basic and clinical science researchers who have contributed in various ways, including providing evidence-informed responses to my questions about whether or not they considered the Streitberger sham needle placed at true acupuncture points to be an inert control; responding to requests on whether they were aware of any relevant published, ongoing or unpublished trials; providing trials for the CAM trials database described in Chapter 3; reviewing and/or commenting on the chapters before or after publication; and/or providing relevant suggestions and advice on any aspect of this work: Ting Bao, Barker Bausell, Josephine Briggs, Farrar Celada, Kevin Chen, Margaret Chesney, Delia Chiaramonte, Mike Clarke, Chris D'adamo, Edzard Ernst, Heather Dubnick, Nadine Foster, Chantelle Garritty, Adele Gilpin, Richard Hammerschlag, Adi Haramati, Richard Harris, Julian Higgins, Marcos Hsu, Kirsten Karchmer, Partap Khalsa, Aakanksha Khandelwal, Christine Laine, Joseph Lau, Myeong Soo Lee, Wenlin Lee, Carol Lefebvre, Byungmook Lim, Klaus Linde, Jianping Liu, Tao Liu, Paul Magarelli, Steve McDonald, Nancy Min, David Moher, Susan Moss, Cynthia Mulrow, Richard Nahin, Vitaly Napadow, Elizabeth Kimbrough Pradhan, Lee Hullender Rubin, Ray Rubio, Margaret Sampson, Rosa Schnyer, Caylie See, Beverly Shea, Byung-Cheul Shin, Caroline Smith, Karen Soeken, Kristen

Stafford, Elizabet Stener-Victorin, Prathap Tharyan, Jayne Tierney, Kiichiro Tsutani, Andrew Vickers, Wendy Weber, Adrian White, Peter White, Grant Zhang, and Qi Zhu.

Finally, I would like to especially thank the renowned and respected group of authorities who agreed to take time out of their busy schedules to review my thesis book and to be on the examination panel: Riekie de Vet, Hans Evers, Bart Koes, Rob Scholten, Gerben ter Riet, and Claudia Witt.

About the author

Eric Manheimer was born in 1969 in Boston, Massachusetts. He attended the Johns Hopkins University in Baltimore, graduating with a BA degree in philosophy in 1992. Subsequently he worked as a research assistant in a basic science research laboratory, prior to starting a Master's degree program. In 1995, he enrolled as a Master's student in Epidemiology at the University of Maryland, Baltimore, Department of Epidemiology and Preventive Medicine. During his Master's studies, he undertook summer internships at the US National Institutes of Health, Office of Dietary Supplements, where he assisted in the development of a Congressionally mandated database of information on dietary supplements called Computer Access to Research on Dietary Supplements (CARDS). After graduating with a Master's degree in Epidemiology in 1997, he worked with Professor Kay Dickersin, a pioneer in publication bias, trial registers, and evidence-based healthcare, for five years as the coordinator and methodologist of the Baltimore/New England/US Cochrane Center. He moved to Providence in 2000, where the Baltimore Cochrane Center was relocated and eventually renamed as the US Cochrane Center. At the Cochrane Center, he oversaw multiple projects related to clinical trial identification and registration, including the development of the Cochrane Central Register of Controlled Trials, which is the most comprehensive database of trials in the world. In January 2003, he accepted a position as the administrator and methodologist of the Baltimore-based Cochrane Complementary Medicine Field, convened by Professor Brian Berman, a pioneer in integrative medicine research, education, and clinical care, who co-founded the Cochrane Complementary Medicine Field in 1996. At the Cochrane CAM Field, he prepares systematic reviews with a focus on CAM therapies; conducts research related to the improvement of the methodology of CAM reviews; organizes and leads workshops at multiple international conferences to provide training for systematic reviewers in CAM; and publishes columns in *Global Advances in Health and Medicine* and *Explore* on the Cochrane Collaboration and Cochrane CAM Reviews. At the CAM Field, he also contributes to establishing part of the infrastructure that supports preparation and maintenance of Cochrane reviews of CAM therapies by coordinating and serving on the review panel for a CAM Field bursary scheme; contributing to the development of the CAM Field register of trials; contributing to the development of the CAM Field Topics List of Cochrane CAM-related reviews; and collaborating with other Cochrane entities on the preparation and dissemination of summaries of CAM-related Cochrane Reviews for the lay public. In 2006, he prepared a competitively reviewed NIH grant application which resulted in the award of a 5-year, $2.1 million USD grant to support the activities of the Cochrane CAM Field, from 2007-2012. In 2011, he prepared another competitively reviewed NIH grant application, which resulted in the award of a 5-year, $3 million USD grant to continue the support of the activities of the Cochrane CAM Field. In 2006, Eric Manheimer began working under the supervision of Professor Lex Bouter, in preparing a systematic review of acupuncture for knee osteoarthritis, which was published in the *Annals of Internal Medicine* in 2007. In 2008, he was admitted as a PhD student at the EMGO-Institute for Health and Care Research, VU University Medical Center, under the supervision of Professors Lex Bouter and Daniëlle van der Windt, conducting much of the work described in this thesis.

List of publications

Most of the publications listed below are available in the open domain, and can be accessed by using the 'List of Publications' link on my personal webpage, available at: https://sites.google.com/site/manheimereric/home

Peer-reviewed journal articles

1. Liu JP, Han M, Li XX, Mu YJ, Lewith G, Wang YY, Witt CM, Yang GY, Manheimer E, Snellingen T, Berman B, Gluud C. Prospective registration, bias risk, and outcome-reporting bias in randomised clinical trials of traditional Chinese medicine: an empirical methodological study. *BMJ Open* 2013 Jul 16;3(7):e002968. doi: 10.1136/bmjopen-2013-002968. PubMed Central PMCID: PMC3717464.
2. Wieland LS, Manheimer E, Sampson M, Barnabas JP, Bouter LM, Cho K, Lee MS, Li X, Liu J, Moher D, Okabe T, Pienaar ED, Shin BC, Tharyan P, Tsutani K, van der Windt DA, Berman BM. Bibliometric analysis of the Cochrane Complementary Medicine Field specialized register of controlled trials. *Syst Rev* 2013 July 4;2:51. doi:10.1186/2046-4053-2-51. PubMed Central PMCID: PMC3704678. **[Thesis Chapter 3]**
3. Manheimer E, van der Windt D, Cheng K, Stafford K, Liu J, Tierney J, Lao L, Berman BM, Langenberg P, Bouter LM. The effects of acupuncture on rates of clinical pregnancy among women undergoing in vitro fertilization: a systematic review and meta-analysis. *Hum Reprod Update* 2013 June 27. doi:10.1093/humupd/dmt026. [Epub ahead of print] **[Thesis Chapter 9]**
4. Xu S, Wang L, Cooper E, Zhang M, Manheimer E, Berman B, Shen X, Lao L. Adverse events of acupuncture: a systematic review of case reports. *Evid Based Complement Alternat Med*. 2013 Mar 20;2013:581203. doi: 10.1155/2013/581203. PubMed Central PMCID: PMC3616356.
5. Turner L, Galipeau J, Garritty C, Manheimer E, Wieland LS, Yazdi F, Moher D. An evaluation of epidemiological and reporting characteristics of complementary and alternative medicine (CAM) systematic reviews (SRs). *PLoS ONE* 2013 Jan 18;8(1):e53536. doi:10.1371/journal.pone.0053536. PubMed Central PMCID: PMC3544927.
6. Vickers AJ, Cronin AM, Maschino AC, Lewith G, Macpherson H, Foster NE, Sherman KJ, Witt CM, Linde K; Acupuncture Trialists' Collaboration. Acupuncture for Chronic Pain: Individual Patient Data Meta-analysis. *Arch Intern Med* 2012 Oct 22;172(19):1444-53. doi: 10.1001/archinternmed.2012.3654. PubMed Central PMCID: PMC3658605. [Manheimer E is a member of the Acupuncture Trialists' Collaboration and contributed to this project and the resulting paper.]
7. Chen KW, Berger CC, Manheimer E, Forde D, Magidson J, Dachman L, Lejuez CW. Meditative therapies for reducing anxiety: a systematic review and meta-analysis of randomized controlled trials. *Depress Anxiety* 2012 Jul;29(7):545-62. doi:10.1002/da.21964. PubMed Central PMCID: PMC3718554.
8. Manheimer E, Cheng K, Wieland LS, Min LS, Shen X, Berman BM, Lao L. Acupuncture for treatment of irritable bowel syndrome: systematic review and meta-analysis. *Am J Gastroenterol* 2012 June;107(6):835-847;quiz 848. doi: 10.1038/ajg.2012.66. PubMed Central PMCID: PMC3671917.
9. Manheimer E, Cheng K, Wieland LS, Min LS, Shen X, Berman BM, Lao L. Acupuncture for treatment of irritable bowel syndrome. *Cochrane Database Syst Rev* 2012, Issue 5. Art. No.: CD005111. doi: 10.1002/14651858.CD005111.pub3. PubMed Central PMCID: PMC3718572. [This review was peer-reviewed dually and independently by the *Am J Gastroenterol* and the Cochrane Inflammatory Bowel Diseases and Functional Bowel Disorders Review Group. This Cochrane version is the unabridged review (i.e., which includes all tables and figures).]
10. Witt CM, Manheimer E, Hammerschlag R, Lüdtke R, Lao L, Tunis SR, Berman BM. How well do randomized trials inform decision making: systematic review using comparative effectiveness research measures on acupuncture for back pain. *PLoS ONE* 2012 Feb 28;7(2):e32399. doi: 10.1371/journal.pone.0032399. PubMed Central PMCID: PMC3289651.
11. Dickersin K, Manheimer E, Li T. Surgery for nonarteritic anterior ischemic optic neuropathy. *Cochrane Database Syst Rev* 2012, Issue 1. Art. No.: CD001538. doi: 10.1002/14651858.CD001538.pub3. PubMed Central PMCID: PMC Journal – In Process. Update of: *Cochrane Database Syst Rev* 2006, Issue 1. Art. No.: C001538, and *Cochrane Database Syst Rev* 2000, Issue 2. Art. No.: C001538.
12. Turner LA, Singh K, Garritty C, Tsertsvadze A, Manheimer E, Wieland LS, Galipeau J, Moher D. An evaluation of the completeness of safety reporting in reports of complementary and alternative medicine trials. *BMC Complement Altern Med* 2011 Aug 22;11:67. doi: 10.1186/1472-6882-11-67. PubMed Central PMCID: PMC3176483.

13. Manheimer E. Selecting a control for in vitro fertilization and acupuncture randomized controlled trials (RCTs): how sham controls may unnecessarily complicate the RCT evidence base. *Fertil Steril* 2011 Jun 30;95(8):2456-61. doi: 10.1016/j.fertnstert.2011.04.040. PubMed Central PMCID: PMC3124624. **[This article was prepared as preliminary work for Thesis Chapter 9]**
14. Wieland S*, Manheimer E*, Berman BM. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane Collaboration. *Altern Ther Health Med* 2011 Mar-Apr;17(2):50-9. PubMed Central PMCID: PMC3196853. [*Both authors contributed equally to this work.] **[Thesis Chapter 2]**
15. Vickers AJ, Cronin AM, Maschino AC, Lewith G, Macpherson H, Victor N, Sherman KJ, Witt C, Linde K; Acupuncture Trialists' Collaboration. Individual patient data meta-analysis of acupuncture for chronic pain: protocol of the Acupuncture Trialists' Collaboration. *BMC Trials* 2010 Sep 28;11:90. doi: 10.1186/1745-6215-11-90. PubMed Central PMCID: PMC2955653. [Manheimer E is a member of the Acupuncture Trialists' Collaboration and contributed to this protocol and the 2012 final paper.]
16. Manheimer E, Cheng K, Linde K, Lao L, Yoo J, Wieland S, van der Windt DA, Berman BM, Bouter LM. Acupuncture for peripheral joint osteoarthritis. *Cochrane Database Syst Rev* 2010, Issue 1. Art. No.: CD001977. doi: 10.1002/14651858.CD001977.pub2. PubMed Central PMCID: PMC3169099. **[Thesis Chapter 8]**
17. Manheimer E, Wieland S, Kimbrough E, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for traditional Chinese medicine therapies. *J Altern Complement Med* 2009 Sep;15(9):1001-14. doi: 10.1089/acm.2008.0414. PubMed Central PMCID: PMC2856612. **[Thesis Chapter 4]**
18. Cogo E, Sampson M, Ajiferuke I, Manheimer E, Campbell K, Daniel R, Moher D. Searching for controlled trials of complementary and alternative medicine: a comparison of 15 databases. *Evid Based Complement Alternat Med* 2009 May 25. doi: 10.1093/ecam/nep038. PubMed Central PMCID: PMC3137728.
19. Linde K, Allais G, Brinkhaus B, Manheimer E, Vickers A, White AR. Acupuncture for tension-type headache. *Cochrane Database Syst Rev* 2009, Issue 1. Art. No.: CD007587. doi: 10.1002/14651858.CD007587. PubMed Central PMCID: PMC3099266.
20. Linde K, Allais G, Brinkhaus B, Manheimer E, Vickers A, White AR. Acupuncture for migraine prophylaxis. *Cochrane Database Syst Rev* 2009, Issue 1. Art. No.: CD001218. doi: 10.1002/14651858.CD001218.pub2. PubMed Central PMCID: PMC3099267.
21. Park J, Linde K, Manheimer E, Molsberger A, Sherman K, Smith C, Sung J, Vickers A, Schnyer R. The status and future of acupuncture clinical research. *J Altern Complement Med* 2008 Sep;14(7):871-81. doi: 10.1089/acm.2008.SAR-4. PubMed Central PMCID: PMC3155101.
22. Manheimer E, Zhang G, Udoff F, Haramati A, Langenberg P, Berman BM, Bouter LM. The effects of acupuncture on pregnancy and live birth rates among women undergoing in vitro fertilization: a systematic review and meta-analysis. *BMJ* 2008 Mar;336(7643):545-9. doi: 10.1136/bmj.39471.430451.BE. PubMed Central PMCID: PMC2265327. **[This review was prepared as preliminary work for Thesis Chapter 9]**
23. Ee C, Manheimer E, Pirotta M, White A. Acupuncture for pelvic and back pain in pregnancy: systematic review. *Am J Obstet Gynecol* 2008 Mar;198(3):254-9. doi: 10.1016/j.ajog.2007.11.008.
24. Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for knee osteoarthritis. *Ann Intern Med* 2007 Jun;146(12):868-77. **[This review was prepared as preliminary work for Thesis Chapter 8]**
25. Manheimer E, Lim B, Lao L, Berman B. Acupuncture for knee osteoarthritis--a randomised trial using a novel sham. *Acupunct Med* 2006 Dec;24 Suppl:S7-14.
26. Lim B, Manheimer E, Lao L, Ziea E, Wisniewski J, Liu J, Berman B. Acupuncture for treatment of irritable bowel syndrome. *Cochrane Database Syst Rev* 2006, Issue 4. Art. No.: CD005111. doi: 10.1002/14651858.CD005111.pub2.
27. Manheimer E, Ezzo J, Hadhazy V, Berman B. Published reports of acupuncture trials showed important limitations. *J Clin Epidemiol* 2006 Feb;59(2):107-13. doi: 10.1016/j.jclinepi.2005.06.003. **[Thesis Chapter 6]**

28. Liu JP, Manheimer E, Yang M. Herbal medicines for treating HIV infection and AIDS. *Cochrane Database Syst Rev* 2005, Issue 3. Art. No.: CD003937. doi: 10.1002/14651858.CD003937.pub2.
29. Manheimer E, White A, Berman B, Forys K, Ernst E. Meta-analysis: Acupuncture for low back pain. *Ann Intern Med* 2005 Apr 19;142(8):651-663. Erratum in: *Ann Intern Med*. 2005 Jun 7;142(11):950-1. **[Thesis Chapter 7]**
30. Liu JP, Manheimer E, Shi Y, Gluud C. Chinese herbal medicines for severe acute respiratory syndrome: a systematic review and meta-analysis. *J Altern Complement Med* 2004 Dec;10(6):1041-51.
31. Gagnier JJ, Chrubasik S, Manheimer E. Harpagophytum procumbens for osteoarthritis and low back pain: A systematic review. *BMC Complement Altern Med* 2004 Sep;4(13). doi: 10.1186/1472-6882-4-13. PubMed Central PMCID: PMC520753.
32. Manheimer E, Anderson BJ, Stein MD. Use and assessment of complementary and alternative therapies by intravenous drug users. *Am J Drug Alcohol Abuse* 2003 May;29(2):401-13.
33. Liu J, Manheimer E, Tsutani K, Gluud C. Medicinal herbs for hepatitis C virus infection: a Cochrane hepatobiliary systematic review of randomized trials. *Am J Gastroenterol* 2003 Mar;98(3):538-44.
34. Manheimer E, Anderson D. Survey of public information about ongoing clinical trials funded by industry: Evaluation of completeness and accessibility. *BMJ* 2002 Sept 7;325(7363):528-31.
35. Dickersin K, Manheimer E, Wieland S, Robinson K, Lefebvre C, McDonald S, and the CENTRAL Development Group. Development of the Cochrane Collaboration's CENTRAL register of controlled clinical trials. *Eval Health Prof* 2002 Mar;25(1):38-64.
36. Liu JP, Manheimer E, Tsutani K, Gluud C. Medicinal herbs for hepatitis C virus infection. *Cochrane Database Syst Rev* 2001, Issue 4. Art. No.: CD003183. doi: 10.1002/14651858.CD003183.
37. Manheimer E, Silbergeld EK. Critique of CDC's retreat from recommending universal lead screening for children. *Public Health Rep* 1998 Jan-Feb;113(1):38-46. PubMed Central PMCID: PMC1308366.

Non-peer-reviewed journal articles

1. Manheimer E. Summary of findings tables: presenting the main findings of Cochrane CAM-related reviews in a transparent and simple tabular format. *Global Adv Health Med* 2012 Mar;1(1):90-1.
2. Stener-Victorin E*, Manheimer E*. Commentary on the Cochrane review of acupuncture and assisted conception. *Explore (NY)* 2011 Mar-Apr;7(2):120-133. doi:10.1016/j.explore.2010.12.016. PubMed Central PMCID: PMC3086273. [*Both authors contributed equally to this work.] **[This article was prepared as preliminary work for Thesis Chapter 9]**
3. Manheimer E, Brian B, Vist G, Glenton C. The effect of probiotics on preventing necrotizing enterocolitis in premature babies. *Altern Ther Health Med* 2009 Jul/Aug;15(4):18-20.
4. Manheimer E, Berman B. Producing and disseminating systematic reviews: A summary of the CAM-related work presented at the 13th International Cochrane Colloquium. *J Altern Complement Med* 2005 Mar;12(2):193-6. doi: 10.1089/acm.2006.12.193.
5. Manheimer E, Berman B. The 13th International Cochrane Colloquium in Melbourne: An invitation for CAM researchers and consumers. *J Altern Complement Med* 2005 Aug;11(4):733-4. doi: 10.1089/acm.2005.11.733.
6. Manheimer E, Berman B. Exploring, evaluating and applying the results of systematic reviews of CAM therapies. *Explore* 2005 May;1(3):210-14. doi: 10.1016/j.explore.2005.02.005.
7. Manheimer E, Berman B. NCCAM Support for Cochrane CAM Field. *Complement Ther Med* 2003 Dec;11(4):268-71.
8. Manheimer E, Berman B. Cochrane for CAM Providers: Evidence for Action. *Altern Ther Health Med* 2003 Sept/Oct;9(5):110-12.
9. Dickersin K, Manheimer E. The Cochrane Collaboration: Evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin Obstet Gynecol* 1998 June;41(2):315-31.

Book chapters

1. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011] The Cochrane Collaboration, 2011.

2. Manheimer E*, Ezzo J*. Chapter 2: Systematic reviews and CAM. In: Adams J (editor). *Researching Complementary and Alternative Medicine*. Oxford: Routledge, 2006, pgs 19-36. [*Both authors contributed equally to this work.] **[Thesis Chapter 5]**
3. Committee on the Use of Complementary and Alternative Medicine by the American Public. Chapter 5: State of emerging evidence on CAM. In: Board of Health Promotion and Disease Prevention, Institute of Medicine of the National Academies. *Complementary and Alternative Medicine in the United States*. Washington DC: The National Academies Press; 2005. (Work was acknowledged by the following statement: "The committee would also like to thank Eric Manheimer for his contributions about emerging evidence in Chapter 5." p. XiV.)

Other brief communications

1. Manheimer E, Berman BM. Letter to the Editor in response to: [Ernst E, Lee MS, Choi TY. Acupuncture: does it alleviate pain and are there serious risks? A review of reviews. *Pain* 2011;152:755-764.] *Pain* 2011 Sept;152(9):2179-80. doi: 10.1016/j.pain.2011.04.009. [author reply 2184-6]
2. Manheimer E, Bouter LM. Does acupuncture improve the results of in vitro fertilization? [letter: response to commentary by Ernst E] *Focus on Alternative Therapies (FACT)* 2008 Sept;13(3):195. **[This brief communication was prepared as preliminary work for Thesis Chapter 9]**
3. Manheimer E, Berman BM, Bouter LM. Effects of acupuncture on rates of pregnancy and live birth among women undergoing in vitro fertilization: systematic review and meta-analysis [letter: response to *BMJ* rapid responses] *BMJ* [online only] 2008 May 2. **[This response to *BMJ* Rapid Responses was prepared as preliminary work for Thesis Chapter 9]**
4. Manheimer E, White A, Ernst E, Langenberg P. Acupuncture for low back pain [letter: response]. *Ann Intern Med* 2005;143(9):692. **[This brief communication was the authors' response to a Letter to the Editor on Thesis Chapter 7]**
5. Gagnier JJ, Chrubasik S, Manheimer E. Harpagophytum for musculoskeletal pain: a systematic review. *Focus on Alternative and Complementary Therapies (FACT)* 2003;8:498.
6. Manheimer E. How the Cochrane controlled trials register is produced. In: *The Cochrane Library*. Oxford: Update Software, 2002. (published electronically in *The Cochrane Library* up until 2002)
7. Lefebvre C, Lusher A, Dickersin K, Manheimer E. Literature searches [letter]. *Lancet* 2002 Mar 9;359(9309):896.

Systematic review protocol

1. Manheimer E, van der Windt D, Stafford K, Liu J, Tierney J, Lao L, Berman B, Bouter L. Effects of adjuvant acupuncture on pregnancy rates among women undergoing in vitro fertilization: systematic review and meta-analysis. PROSPERO 2011:CRD42011001309. **[This protocol was prepared as preliminary work for Thesis Chapter 9]**

Published multimedia

1. Manheimer E, Wieland S, Kimbrough E, Cheng K, Lao L, Berman B. Evidence from the Cochrane Collaboration for Traditional Chinese Medicine. Natural Standard webinar presented by Eric Manheimer, and published online Sept 2010. [Natural Standard website only allows paid subscribers to view the webinar.]

Abstracts and/or proceedings

1. Sæterdal I, Nilsen ES, Underland V, Glenton C, Vist G, Manheimer E. Preparation of summaries of Cochrane CAM reviews for consumers. In: 5th International Congress on Complementary Medicine Research; 2010 May 19-21; Tromso, Norway. p. 67-68. Abstract no. P-022. [This poster was awarded one of the three poster prizes.]
2. Wieland S, Manheimer E. Issues in improving the quality and increasing the size of the Cochrane Complementary and Alternative Medicine (CAM) Field specialized register. In: 17th International Cochrane Colloquium; 2009 Oct 11-14; Singapore. p. 47-48. Abstract no. P56. **[This abstract was prepared as preliminary work for Thesis Chapter 3]**

3. Manheimer E, Kimbrough E, Wieland S, Cheng K, Berman BM. Evidence from the Cochrane Collaboration for traditional Chinese medicine therapies. In: 17th International Cochrane Colloquium; 2009 Oct 11-14; Singapore. p. 58-59. Abstract no. P85. **[This abstract was prepared as preliminary work for Thesis Chapter 4]**
4. Manheimer E, Berman B, Dubnick H, Beckner W. Cochrane reviews of complementary and alternative therapies: evaluating the strength of the evidence. In: 12th International Cochrane Colloquium; 2004 Oct 2-6; Ottawa. p. 159-160. Abstract no. P-094. **[This abstract was prepared as preliminary work for Thesis Chapter 4]**
5. Liu JP, Manheimer E, Yang M. Herbal medicines for treating HIV infection and AIDS: a Cochrane systematic review. In: XV International AIDS Conference; 2004 July 11-16; Bangkok. p. 268.
6. Berman B, Manheimer E, Beckner M. Building the framework for a collaborative, international, evidence-based approach to the evaluation of Traditional Chinese Medicine [selected for oral presentation for: Session 13: Evidence and international co-operation in developing countries]. In: 11th International Cochrane Colloquium: evidence, health care and culture; 2003 Oct 26-31; Barcelona. p. 25. Abstract no. O-60.
7. Wieland S, Manheimer E, Brodney-Folse S. Status of PubMed records classified as CCT on the basis of abstracts. In: 11th International Cochrane Colloquium: Evidence, health care and culture, 2003 Oct 26-31; Barcelona. p. 78. Abstract no. P-148.
8. Manheimer E, Lefebvre C, Timimi H, Rutks I, Ghersi D. Lessons learned and ongoing challenges in developing an international, collaborative register of controlled trials with many contributors. In: 10th International Cochrane Colloquium; 2002 Jul 31 –Aug 3; Stavanger. p. 26. Abstract no. P14.
9. Liu J, Manheimer E, Tsutani K, Gluud C. Medicinal herbs for hepatitis C: a systematic review of randomized trials. In: 9th International Cochrane Colloquium: The evidence dissemination process – how to make it more efficient; 2001 Oct 9-13; Lyon. p. 80. Abstract no. P-097; Session B.
10. Manheimer E, Wieland S, Dickersin K. The current composition of the Cochrane Collaboration's CENTRAL database of controlled trials of healthcare interventions. In: 9th International Cochrane Colloquium: The evidence dissemination process – how to make it more efficient; 2001 Oct 9-13; Lyon. p. 76. Abstract no. P-088; Session B.
11. Wieland S, Manheimer E, Dickersin K. Results of reviewing MEDLINE records without abstracts. In: 9th International Cochrane Colloquium: The evidence dissemination process – how to make it more efficient; 2001 Oct 9-13; Lyon. p. 75. Abstract no. P-087, Session B.
12. Manheimer E, Anderson D, Dickersin K. Identifying ongoing clinical trials online: The need for a comprehensible, comprehensive register. 22nd Annual Meeting of the Society for Clinical Trials; 2001 May 20-23; Denver. [Abstract published in: *Control Clin Trials* 2001;22(2S):100S].
13. Manheimer E, Wieland S, Robinson K, Lefebvre C, McDonald S, Clarke M, Starr M, Dickersin K. The Cochrane Collaboration's CENTRAL register of controlled trials: Development and current contents. 22nd Annual Meeting of the Society for Clinical Trials; 2001 May 20-23; Denver. [Abstract published in: *Control Clin Trials* 2001;22(2S):40S].
14. Manheimer E, Anderson D, Dickersin K. Dissemination of information about ongoing clinical trials: Can we find a better approach? In: 8th International Cochrane Colloquium: Evidence for Action – Challenges for the Cochrane Collaboration in the 21st Century; 2000 Oct 25-29; Cape Town. p. 190. Abstract no. PD90.
15. Manheimer E, Fedder DO. A church-based lead education program. In: National Conference on Health Education and Health Promotion and the Society for Public Health Education Midyear Scientific Conference; 1997 May 30; Atlanta GA. Abstract no. A27.
16. Wang SM, Manheimer E, Darlington DP. Salt-induced rennin-like mRNA in Dahl rats. *Am J Hypertens* 1995;8(4), Part 2:49A. Abstract no. A6.

Major invited speeches

National

1. Manheimer E. Acupuncture for low back pain -- Meta-analysis of efficacy. [Session: CAM modalities]. Presented at: Johns Hopkins Medicine Continuing Medical Education Conference: Integration of Complementary and Alternative Medicine into Clinical Practice: Evidence-based Medicine; 2005 Jun 8-10; Baltimore.
2. Sieving P, Manheimer E. Hands-on: Getting the most from the Cochrane Library [workshop]. Workshop presented at: US Cochrane Collaboration Meeting: Creating Greater Awareness and Use of Evidence-based Health Care; 2004 Apr 1-2; Providence. Workshop 3.

International

1. Berman B, Manheimer E. Navigating to knowledge: the Cochrane Collaboration Complementary Medicine Field [Joint keynote presentation]. Presented at: 8th International Congress on Complementary Medicine Research; 2013 April 11-13; London, UK.
2. Manheimer E, Wieland LS, Liu JP. Systematic review research methods training: Focus on CAM [workshop]. Pre-conference workshop at: 8th International Congress on Complementary Medicine Research; 2013 April 11-13; London, UK.
3. Wieland LS, Manheimer E. Systematic reviews and meta-analyses [workshop]. Invited by the President of the International Society for Complementary Medicine (ISCMR) to present a 2-day training workshop at: ISCMR's 7th Research Methods in Complementary and Integrative Medicine training course; 2013 March 23-24; Baltimore, Maryland.
4. Manheimer E, Wieland LS, Liu JP. Systematic review research methods training: Focus on CAM [workshop]. Pre-conference workshop at: 2012 International Research Congress on Integrative Medicine & Health/7th International Congress on Complementary Medicine Research; 2012 May 15-18; Portland, Oregon.
5. Manheimer E, Wieland LS. Systematic reviews and meta-analyses [workshop]. Invited by the President of the International Society for Complementary Medicine (ISCMR) to present a 2-day training workshop at: ISCMR's 6th Research Methods in Complementary and Integrative Medicine training course; 2012 March 29-April 4; Baltimore, Maryland.
6. Manheimer E, Wieland LS, Liu JP. How to prepare and finalize a Cochrane systematic review on traditional Chinese medicine therapies [workshop]. Workshop presented at: 6th International Congress on Complementary Medicine Research; 2011 May 6-9; Chengdu, China. p. 11.
7. Thomson D, Wieland LS, Stern C, Seto I, Pearson A, Manheimer E. Practical aspects of developing and using Fields topics lists in Archie [workshop]. Workshop presented at: 18th International Cochrane Colloquium: Joint Colloquium of the Cochrane and Campbell Collaborations; 2010 Oct 18-22; Keystone, Colorado. p. 143. Abstract no. WS23.
8. Wieland S, Manheimer E. Introduction to systematic reviews and meta-analysis: Focus on CAM [workshop]. Workshop presented at: 5th International Congress on Complementary Medicine Research; 2010 May 19-21; Tromso, Norway.
9. Manheimer E (Chair). Conducting research on Traditional Chinese Medicine Therapies [Special session 5]. In: 17th International Cochrane Colloquium; 2009 Oct 11-14; Singapore. p. 56.
10. Manheimer E, Horneber M. Co-Chairs. Invited session 2: Can we trust systematic reviews of complementary therapies? 16th International Cochrane Colloquium; 2008 Oct 3-7; Freiburg, Germany. p. 58.
11. Manheimer E, Wieland LS. Use of field topic lists in Archie [workshop]. Workshop presented at: 16th International Cochrane Colloquium; 2008 Oct 3-7; Freiburg. p. 92. Workshop W36.
12. Manheimer E. A systematic review and meta-analysis of acupuncture for knee osteoarthritis. Presented at: 2007 Annual Conference of the Society for Acupuncture Research. 2007 Nov 9-11; Baltimore.
13. Manheimer E. Acupuncture for irritable bowel syndrome – a systematic review. Presented at: International Symposium on Traditional Chinese Medicine Approaches to Gastrointestinal Disorders. Organized by the Society for Acupuncture Research. 2007 Nov 8; Baltimore.

14. Manheimer E. Exploring the challenges of complementary medicine reviews [workshop]. Workshop presented at: 15th International Cochrane Colloquium; 2007 Oct 23-27; Sao Paulo. p. 209. Workshop W48.
15. Manheimer E, Cahill K, Wieland S, Becker L. Use of field topic lists in Archie [workshop]. Workshop presented at: 15th International Cochrane Colloquium; 2007 Oct 23-27; Sao Paulo. p. 195. Workshop W20.
16. Manheimer E. Systematic reviews [workshop]. Workshop presented at The Galilee Society – The Arab National Society for Health Research and Services; 2007 Apr 16; Shefa Amr.
17. Manheimer E, Haramati A, Berman BM. Introduction to systematic reviews and meta-analysis [workshop]. Workshop presented at 18th Israeli Medical Association World Fellowship International Conference; 2007 Apr 11-15; Jerusalem.
18. Manheimer E. Acupuncture for knee osteoarthritis: a randomized trial using a novel sham. Presented at: International Symposium on Evidence-based Acupuncture Now. Organized by the Japan Society for Acupuncture and Moxibustion. 2006 Nov 20-21; Kyoto.
19. Manheimer E, Liu J. Methodological issues in systematic reviews of complementary and alternative medicine therapies [workshop]. Workshop presented at: 14th International Cochrane Colloquium; 2006 Oct 23-26; Dublin. p. 33. Workshop W49.
20. Manheimer E, Nedrow A. Methodological challenges of preparing systematic reviews of randomized controlled trials in CAM [workshop]. Workshop presented at: North American Research Conference on Complementary and Integrative Medicine; 2006 May 24-27; Edmonton.
21. Manheimer E, Liu J. Issues in preparing systematic reviews of complementary and alternative medicine therapies [workshop]. Workshop presented at: 13th International Cochrane Colloquium; 2005 Oct 22-26; Melbourne. p. 181. Published Abstract no. W22.
22. Manheimer E, Higgins G, Wieland S. Using ProCite for creating and submitting trials databases to CENTRAL [workshop]. Workshop presented at: 13th International Cochrane Colloquium; 2005 Oct 22-26; Melbourne. p. 173. Published Abstract no. W06.
23. Lao L, Manheimer E, Berman B. Systematic reviews of Chinese Medicine. Presented at: 2005 International Society for Chinese Medicine: International Conference on Chinese Medicine: Chinese Medicine for the Prevention and Treatment of Contemporary Diseases, An Evidence-based Approach; 2005 Mar 31 – Apr 3; Macao. p. 27-28. Published Abstract.
24. Manheimer E, White A, Berman B, Forys K, Ernst E. Acupuncture for the treatment of low back pain: evidence from a meta-analysis of randomized controlled trials. Presented at: 2005 International Society for Chinese Medicine: International Conference on Chinese Medicine: Chinese Medicine for the Prevention and Treatment of Contemporary Diseases, An Evidence-based Approach; 2005 Mar 31 – Apr 3; Macao. p. 39-40. Published Abstract.
25. Manheimer E. Finding all appropriate studies. Presented at: Systematic Reviews in Dental and Craniofacial Research Symposium [Seq#:293]. 83rd General Session and Exhibition of the International Association for Dental Research; 2005 Mar 12; Baltimore. p. 162. Presentation No. 2647.
26. Manheimer E, Jianping L, Berman B, Gagnier J. Tackling the challenges of preparing systematic reviews of complementary and alternative medicine interventions: acupuncture as an example [workshop]. Workshop presented at: 12th International Cochrane Colloquium; 2004 Oct 2-6; Ottawa. Workshop W-027.
27. Glatman E, Wieland S, Manheimer E. Identifying controlled trial reports (advanced) [workshop]. Workshop presented at: 12th International Cochrane Colloquium; 2004 Oct 2-6; Ottawa. Workshop W-074.
28. Glatman E, Wieland S, Manheimer E. Identifying controlled trial reports (introductory) [workshop]. Workshop presented at: 12th International Cochrane Colloquium; 2004 Oct 2-6; Ottawa. Workshop W-043.
29. Manheimer E, Berman B, Beckner W, Liu J, Gagnier J. Methodological challenges of preparing systematic reviews of complementary and alternative therapies [workshop]. Workshop presented at: 11th International Cochrane Colloquium: Evidence, health care and culture; 2003 Oct 26-31; Barcelona. Workshop 63.

30. Wieland S, Manheimer E, Bluemle A. Advanced handsearching - Identifying and classifying reports of controlled trials in healthcare journals [workshop]. Workshop presented at: 11th International Cochrane Colloquium: Evidence, health care and culture; 2003 Oct 26-31; Barcelona. Workshop 1.
31. Manheimer E, Wieland S, Pienaar E. Creating trials databases for submission to CENTRAL using ProCite reference management software [workshop]. Workshop presented at: 10th International Cochrane Colloquium; 2002 Jul 30 – Aug 3; Stavanger.
32. Wieland S, Manheimer E, Rutks I. Determining study design classification of potential trialreports using MEDLINE abstracts as examples (advanced) [workshop]. Workshop presented at: 10th International Cochrane Colloquium; 2002 Jul 30 – Aug 3; Stavanger.
33. Wieland S, Manheimer E, Rutks I, Dunn, K. Determining study design classification of potential trial reports using MEDLINE abstracts as examples (introductory) [workshop]. Workshop presented at: 10th International Cochrane Colloquium; 2002 Jul 30 – Aug 3; Stavanger.
34. Manheimer E, Wieland S, McGuire H. Using ProCite for creating and maintaining trials registers for submission to CENTRAL [workshop]. Workshop presented at: 9th International Cochrane Colloquium: The evidence dissemination process – how to make it more efficient; 2001 Oct 9-13; Lyon. p. 129. Published Abstract no. C3.
35. Irvin E, Garritty C, Wieland S, Manheimer E. Handsearching the healthcare literature to identify randomized controlled trials and controlled clinical trials (introductory) [workshop]. Workshop presented at: 9th International Cochrane Colloquium: The evidence dissemination process – how to make it more efficient; 2001 Oct 9-13; Lyon. p. 125. Published Abstract no. B5
36. Wieland S, Manheimer E. Handsearching the healthcare literature to identify randomized controlled trials and controlled clinical trials (advanced) [workshop]. Workshop presented at: 9th International Cochrane Colloquium: The evidence dissemination process – how to make it more efficient; 2001 Oct 9-13; Lyon. p. 122. Published Abstract no. A13
37. Manheimer E. Learning to identify and classify reports of controlled trials in healthcare journals (introductory) [workshop]. Workshop presented at: 8th International Cochrane Colloquium: Evidence for Action – Challenges for the Cochrane Collaboration in the 21st Century; 2000 Oct 25-29; Cape Town. p. 236.
38. Manheimer E. Learning to identify and classify reports of controlled trials in healthcare journals (advanced) [workshop]. Workshop presented at: 8th International Cochrane Colloquium: Evidence for Action – Challenges for the Cochrane Collaboration in the 21st Century; 2000 Oct 25-29; Cape Town. p. 230.
39. Manheimer E, Downing-Park J. The handsearch process: From trial identification to submission (beginners) [workshop]. Workshop presented at: 7th International Cochrane Colloquium: The Best Evidence for Health Care – The Role of the Cochrane Collaboration; 1999 Oct 5-9; Rome. p. 106. Published Abstract no. 46.
40. Manheimer E, Downing-Park J. The handsearch process: From trial identification to submission (advanced) [workshop]. Workshop presented at: 7th International Cochrane Colloquium: The Best Evidence for Health Care – The Role of the Cochrane Collaboration; 1999 Oct 5-9; Rome. p. 97. Published Abstract no. 30.
41. Downing-Park J, Manheimer E. Helping handsearchers distinguish randomized controlled trials and controlled clinical trials from other types of study reports: Examples from the literature [workshop]. Workshop presented at: 7th International Cochrane Colloquium: The Best Evidence for Health Care – The Role of the Cochrane Collaboration; 1999 Oct 5-9; Rome. p. 82. Published Abstract no. 6.
42. Manheimer E. The MEDLINE retagging project and its importance to CENTRAL/CCTR. How can Review Groups Help? Presented at: 6th International Cochrane Colloquium: Systematic reviews – Evidence for Action. Annual Meeting of the Cochrane Review Group Coordinators and Trials Search Coordinators; 1998 Oct 21; Baltimore.