REVIEW

CNS Neuroscience & Therapeutics

Acupuncture for Depression: A Critique of the Evidence Base

Sylvia Schroer & Joy Adamson

Health Sciences Department, ARRC Building, University of York, Heslington, York YO10 5DD, UK

Keywords
Acupuncture; Complementary therapies; Complex interventions; Depression.

Correspondence
Joy Adamson, Department of Health Sciences, ARRC Building, Heslington, York YO10 5DD, UK.
Tel.: +44 (0)1904 321378;
Fax: +44 (0)1904 321387;
E-mail: ja14@york.ac.uk

doi: 10.1111/j.1755-5949.2010.00159.x

SUMMARY

Aim: The aim of this review is to examine the evidence for acupuncture's effectiveness as a depression intervention. Unlike other reviews, which consider methodological concerns relevant to all experimental evaluations, this review focuses on the scope of studies, and uses a PICO (patients, intervention, comparison, and outcome) structure to determine what can potentially be learned from primary studies that have already been screened for methodological quality by reviewers. **Discussion:** The review identified a number of study limitations. (i) *Patients*: majority of trial reports have not described a rationale for the selection of patients or inclusion/exclusion criteria. Prognostic indicators were not reported and there were also concerns about the generalizability of study populations. (ii) *Intervention*: most trials investigate poorly rationalized standardized acupuncture protocols thus quality of care may be an issue and generalizability to routine clinical practice is a main concern. In trials using other methods generalizability is also poor. (iii) *Comparisons*: concerns were raised about using therapeutically inappropriate acupuncture. (iv) *Outcomes*: short-term focus and the narrow range of outcomes explored. According to more recent systematic review evidence it is probable the shortcomings identified in the PICO review have not been addressed by subsequent research. The concept of model validity, proposed by other researchers, is discussed, and suggestions put forward about complex intervention evaluation methods, which may be better suited to evaluating acupuncture care. **Conclusion:** Uncertainty remains about the value of acupuncture care, as it is routinely practiced in the West, and this uncertainty has not been resolved by trials to date. Existing evaluations may however be useful for guiding decisions about the value of specific techniques for patients with depression.

Introduction

Depression is a major problem for the world today. Unipolar major depression accounts for 4.4% of the world's total burden of disease as measured by Disability Adjusted Life Years (DALYs).[1] Worldwide, depression is the leading cause of years lost due to disability for incident cases (YLD's), and the fourth leading cause of disability overall (DALYs). Based on current trends the outlook for the future looks even bleaker and depression will become the second leading cause of lost DALYs by 2020 [see Ref. 1, pp. 250–251]. The economic burden of this illness on society is huge, although only a small proportion of the total cost is due to treatment. For example, a UK study identified total costs of £9 billion with £370 million of this figure representing direct treatment costs [2]. Thus the majority of the economic burden of depression comes from indirect costs including lost working days and mortality (premature death due to suicide), which increase if the illness is not treated [3]. However, a significant proportion (50–60%) of patients with depression do not fully

[1]The DALY is a health gap measure introduced in 1993 by the Harvard School of Public Health in collaboration with the World Bank and The World Health Organisation to assess the Global Burden of Disease (GBD). DALYS are essentially a summation of years of life lost due to premature mortality (YLL), years of life lost due to disability for incident cases (YLD) and potential years of life lost due to premature death or less than full health (PYLL).

respond to medication [4,5]. A further problem is that psychological treatments, for example cognitive behavior therapy, for which there is good evidence of effectiveness for depression [6], may not be suitable for a number of patients with depression if cognitive functioning is too impaired by illness for the intervention to be feasible. Doctors acknowledge depression as an area of clinical practice where there is an effectiveness gap [7] whilst patients have expressed a desire to see more options [6], and more effective solutions [8]. It is thus of paramount importance that new treatments for depression are identified and it is this context, that it is of interest that when acupuncture, which originated in China, was introduced to the West, a main focus for therapeutic work became psychological and emotional concerns of patients [9]. Indeed, the primary reason for using acupuncture is, according to a survey conducted in the USA, for "mood care" [10]. It is possible that acupuncture has something to offer. The purpose, or aim of this review is to examine the evidence for acupuncture's effectiveness as a depression intervention, and outline what can potentially be learned about the benefits, or disadvantages, of acupuncture care for depression, from the trials and reviews to date.

When reviewing literature it is necessary to have an understanding of the etiology of the illness under investigation, its natural course, and risk factors that may potentially limit an individual's capacity for recovery. Trialists may need to monitor such factors, as their distribution may not be equal across different intervention groups. In this context, because it relates to prognosis for recovery, it is important to highlight the fact that depression is increasingly coming to be considered as a chronic episodic condition rather than an acute once in a lifetime problem [11]. Individuals with three or more episodes have relapse rates of 70–80% within 3 years but those with nor previous record of depression have only a 20–30% relapse rate over a similar time period [12,13]. Segal et al. suggest that people are at a critical juncture at their first episode of depression. Early onset (before 20 years of age) also seems to be associated with poorer long-term prognosis and higher relapse rates [14]. A review of the epidemiology of depression in primary care [15] has concluded that initial severity of symptoms and presence of a comorbid medical condition were key predictors of persistence of depression. In view of the chronic, episodic nature of depression, researchers conclude that studies of depression interventions should investigate both short term and long-term health outcomes [16].

Before discussing the research and its findings, it may be useful to provide the reader, who may be unfamiliar with acupuncture practice, with some basic information about its characteristic defining features. Popular notions that this is an ancient treatment and that it exists now as it has for thousands of years need to be dispelled. Acupuncture is "the youngest of all the therapeutic modes introduced into the Chinese spectrum of therapies" [17]. Unschuld has traced Chinese medical thinking from its earliest records through the millennia, explaining how it has been influenced and shaped by successive belief systems that include demonological ideas, Confucianism, Taoism, Buddhism, and finally Marxism. In contemporary China and the West, Chinese medicine is characterized by a diversity and plurality of approaches [18–20]. Although traditionalist Chinese medical acupuncture approaches are diverse in their form, some concepts and ideas remain as core features. Two of these, the concepts of "Qi" and "intention," are discussed further here. As will come to light in the review it is apparent that both these aspects of the intervention have been overlooked in experimental investigations of acupuncture. Possibly because both pose a conceptual challenge, their existence has simply been ignored. The intangible nature of Qi has made it problematic to define and it is considered "the most fundamental, misunderstood and controversial concept in traditional Chinese medicine (TCM)" [21]. As a result of difficulties in finding an appropriate way of translating the concept of Qi it has often "loosely" been mistranslated as "energy," a term that conveys "plausibility" but "limits scientific inquiry and biases research" (from Ref. 21). Vitalistic ideas about Qi are considered problematic because they have perpetuated the notion that acupuncture, being founded on metaphysical concepts, cannot be investigated by modern scientific methods. Schnyer et al. suggest that Qi is best viewed not as some sort of vitalistic force but as a "model or framework for organizing correlations, a theory for predicting observations." Birch and Felt discuss Qi not as an entity but as a dynamic associated with "movement and processes" [20, p. 100]. These authors suggest the Qi "paradigm" involves the idea that everything in nature and in the human body is in a constant state of flux according to regularly observable cycles, and highlight the fact that there are many different types of Qi identified by Chinese medical theory [20, pp. 100–105). Intention is another fundamentally important Chinese medical concept that is somewhat difficult to define. Scheid and Bensky use the term significarion, rather than intention to translate the Chinese medical concept of "yi," and criticize romanticist interpretations that have become attached to the concept [22]. Intention is discussed in Chinese medical classics by physician scholars such as *Sun Simiao* and *Zhu Danxi*. In pre-Han texts it is conceived as: "the meditation-like bodily practice of 'intending' whereby one opens oneself to the universal flow of *Qi* so as to acquire a luminous awareness of the world"

[22]. A modern interpretation of intention is outlined by Hicks [23], who is, like Bensky, a renowned educator of Chinese medicine practice in the West, and who considers intention to be a factor that takes practitioners "from the level of technical expertise into the realm of healing." Hick's clearly believes that intention can make a discernable difference to treatment outcome and describes its two main ingredients as [1] the ability to "do nothing," relax and let the result emerge by itself and [2] having a clear outcome in mind. There is a strong sense of "resting inside" for the practitioner and of allowing a result to happen [23].

Another defining feature of TCM acupuncture is that the treatment is usually repeated over time, with appropriate modifications leading researchers to consider the role of learning as being central to its mechanism of action. For example, Bensoussan suggests that acupuncture is best conceived as a type of educational intervention that involves a process of physiological and bodily learning [24]. This, he argues, is a more congruent way of describing acupuncture's mechanism of action than other commonly used explanations such as the "gate control" theory of pain relief (proposed by Melzack and Wall during the early 1960s).

It is also important to recognize that traditional acupuncture, based on Chinese medical theories, is a highly iterative treatment process whereby the patient and practitioner are involved in a close relationship involving verbal and nonverbal communication processes. Research on acupuncture has suggested that aspects of treatment that might normally be considered incidental to its success or failure are in fact characteristic features of the intervention because they are informed by Chinese medical theories [25]. For example, the ways in which a practitioner considers and discusses the patient's medical history and interacts with the patient are unique to Chinese medicine consultations and differ from conventional medicine [25]. The practitioner palpates and touches the body in highly specific ways and practitioner's senses, even the sense of smell, might potentially play a vital role in diagnosis and clinical decision making. Diagnosis and treatment are individualized and the duration of treatment is likely to vary from person to person, depending on responsiveness to the intervention.

The aim of the previous discussion has been to convey the to the reader that Chinese medical acupuncture is a complex intervention that cannot simply be characterized by the mere act of inserting needles into people in prescribed points according to whatever illness they present with. In fact, the location of acupuncture points, and needle stimulation methods vary with different traditions and approaches. Insertion of needles may not even take place with some methods, for example, Toyohari, a form of Meridian Therapy acupuncture that has been practiced in Japan and more recently exported to Europe, USA, and Australasia, uses needles and other metal implements to stimulate *Qi*, but rarely is the skin actually pierced by a needle. The concept of the "live" or "active" acupuncture point is used within this tradition to potentially obtain optimal results, and point locations sometimes alter following consensus agreement between senior practitioners within the Toyohari organization. The idea that acupuncture practice can be defined by the insertion of needles into rigidly fixed point locations, which have remained the same for thousands of years, is simply incorrect. Having described some key factors relating to depression and discussed some characteristic defining features of traditional Chinese medical acupuncture, the effectiveness evidence is evaluated below with an overview of trials to date and systematic reviews.

Study Selection

Study selection was based on those selected for inclusion in a Cochrane review, with the addition of two further studies [26,27]—because the research team responsible for the highest quality trial in the Cochrane review [28] had been conducting further research, and it was decided to include their subsequent studies in the review.

The aim of this review, which was conducted for a PhD thesis was to ascertain what could potentially be learned from the existing evidence base about acupuncture for depression, in order to inform the design of new study/pilot trial. The focus of the review was on the scope of the studies, and whether there were knowledge gaps—because studies had simply not been designed to answer certain questions. In view of this, the review was structured around the four crucial elements of a good clinical research question, described by the acronym PICO (patients, intervention, comparison, and outcome), suggested by Sackett et al. [29]. These elements are the quintessential components of a randomized controlled trial (RCT), when viewed as an experiment that measures the consequences of different forms of treatment on a particular group of individuals with a defined health problem.

A literature search was conducted to identify more recent primary studies, via systematic reviews, in order to update the manuscript for publication in this journal. Four more recently conducted systematic reviews were identified by the search [30–33]. Whilst many more trials have been conducted in China since this original review was written, with the most recent review identifying 20 trials of acupuncture for major depressive disorder for inclusion in a meta analysis [33], it was felt that the arguments put forward, and the issues raised by the critique,

Table 1 Primary studies included in the review

First author and year	Interventions	Results
Luo et al. 1985 (China)	Electro acupuncture vs. Amitriptyline	Both groups significantly improved, no significant difference between them
Luo et al. 1988 (China)	Electro acupuncture vs. Amitriptyline	No significant difference between EA and Amitriptyline on primary outcome measure. Biochemical markers were different for intervention and control
Xiujuan et al. 1994 (China)	Combined electro acupuncture and manual point stimulation—four TCM patterns vs. Amitriptyline	Both groups significantly improved, no significant difference between them except for 'anxiety somatization' which was significantly better in the acupuncture group
Luo et al. 1998 (China)	Two study phases. Phase 1: Electro acupuncture plus placebo medication; amitriptyline; electro acupuncture plus amitriptyline. Phase 2: two groups, EA plus placebo or amitriptyline (Only phase 1 was reported in Smith and Hay Cochrane review, phase 2 appears to be a duplicate of Luo 1988)	Phase 1 – no significant difference between groups. Phase 2 – no significant differences between groups on primary outcome measure but EA had less side effects. Factor analysis showed EA better for anxiety somatization, cognitive disturbance and treatment of reactive depression than amitriptyline. Biochemical study showed plasma nor epinephrine changed greatly after EA treatment.
Allen et al. 1998 (USA)	Manualized specific acupuncture (SPEC) vs. nonspecific active acupuncture comparison (NSPEC) vs. wait list	SPEC significantly better than NSPEC but not significantly better than wait list control
Roschke et al. 2000 (Germany)	Standardized whole body acupuncture plus Mianserin vs. placebo acupuncture plus Mianserin vs. Mianserin	Patients receiving both types of acupuncture improved slightly more than those receiving Mianserin only
Han et al. 2002 (China)	Electro acupuncture vs. Maprotiline	Both groups significantly improved, no significant difference between them
Manber et al. 2004 (USA)	Manualized specific acupuncture (SPEC) vs. nonspecific active acupuncture comparison (NSPEC) vs. Massage	SPEC acupuncture was better than NSPEC but difference not significant. Both were significantly better than massage
Allen et al. 2006 (USA)	Manualized specific acupuncture (SPEC) vs. nonspecific active acupuncture comparison (NSPEC) vs. wait list	No significant difference for SPEC and NSPEC but both significantly better than wait list control.

still hold strong. This is because all of the newer studies have evaluated protocol acupuncture. The advantages and drawbacks of this type of evaluation are covered in the discussion on interventions. Newer evidence has the same limitations as the primary studies described in this review and it was decided to present the review in its original format.

Review Findings

The critique of trials is organized around PICO headings [34]. Primary studies can be seen in Table 1, which shows the interventions and trial results. Having obtained a translation of the Luo et al.'s 1988 trial [35], it appears that phase two of the Luo et al. study [36] is a duplicate of the 1988 trial. It is impossible to be certain about this however as there is no mention in the 1998 trial report of when the research took place, but the details of the two trials are essentially the same.

The Study Population

Study populations or patient groups are summarized in Table 2.

Depressed patients are a heterogeneous group and it may be that acupuncture is more, or less effective, or more or less acceptable, to certain subgroups, within the depressed population as a whole. It would therefore be helpful if trialists offered a rationale for the development and selection of inclusion/exclusion criteria for their study. With the exception of one trial that offered a clear rationale justifying the selection of pregnant women as a trial focus, this was not done. Manber et al. [26] focused on pregnant women, citing the fact that the treatment of depression during pregnancy has been prioritized as an area for clinical improvement and management and

Table 2 Patient groups used in trials of acupuncture for depression

Trial (first author, date)	Patient groups
1. Luo et al. 1985 (China)	47 men and women, who scored over 20 on the Hamilton Rating Scale
2. Luo et al. 1988 (China)	241 men and women (aged 32–64) recruited from three psychiatric hospitals, who scored over 20 on the Hamilton Rating Scale
3. Xiujuan et al. 1994 (China)	41 men and women with clinical depression (Hamilton Rating Scale of 25+), recruited from both in and out patient clinics at the Beijing Medical University, no exclusion criteria reported. 18 people described as having manic depression, 10 had involutional depression and 13 had depressive neurosis. Duration of disease ranged from 4 months to 5 years
4. Luo et al. 1998 (China)	Two phases of research 29 men and women, who scored over 20 on the Hamilton Rating Scale and were recruited from a closed ward in Beijing Medical University Hospital for a pilot trial. All participants were drug free for the week before commencing the trial. Mean age was 36 years and mean course of depression 7.9 years. In the second phase of research (see Ref. [35]) 241 patients recruited from 10 psychiatric hospitals. 193 were bipolar disorder and 48 with reactive depression, all were drug free for 1 week before trial commenced.
5. Allen et al. 1998 (USA)	38 women aged 18–45, with major depression as described by DSM IV, assessed by SCID interview Exclusion criteria: dysthymia or chronic depression, history of psychosis or mania, substance abuse, any current treatment, endocrine abnormalities, history of central nervous system lesions or any medical condition causing depression, pregnancy, suicide potential.
6. Roschke et al. 2000 (Germany)	70 hospital in-patients aged 20–70 years, in Germany. Diagnosed with clinical depression equating to DSM III-R and a score of greater than 18 on the Hamilton Depression Scale. 43 participants had recurrent depressive disorder (DSM code 296.3), 27 had single episode (DSM code 296.2) Participants excluded if suicidal, diagnosis of schizophrenia or bipolar affective disorders or delusions. Patients with coagulation disease, wound healing disease, emphysematous thorax, abnormal blood cell count, serious liver and kidney disease, epilepsy.
7. Han et al. 2002	66 men and women aged 18–55 were recruited to the trial from Beijing University mental institute, China. Inclusion criteria were ICD 10 and a score of 20 or more on the Hamilton rating Scale.
8. Manber et al. 2004	61 pregnant women with a Hamilton rating score of 14 or more. Participants recruited from obstetric clinics and local magazine advertisements. Inclusion criteria 18 years +, gestational age between 11 and 28 weeks at screening. Exclusion criteria: index MDE of 2 years+; psychotic features; a seasonal pattern; current active suicide potential; cluster B Axis II disorder or other Axis I disorder in past 2 months, except for simple phobia, social phobia or GAD (determined by SCID-IV and SCID-II; abnormal thyroid panel; an uncontrolled medical condition; a condition that may be a medical basis for depression; current use of any medication that impacts mood; confounding treatments for depression; and conditions that require bed rest.
9. Allen et al. 2006	151 participants (104 women, 47 men) who met DSMIV criteria for MDD, assessed by SCID-P, and had a rating score or 14+ on the Hamilton Rating Scale for depression. Exclusion criteria: dysthymia or chronic (>2 yrs) MDD; seasonal pattern; any current Axis 1 diagnosis besides MDD or any Axis II Cluster B disorder; history of psychosis or mania; substance abuse or dependence within the past 4 months; any current relevant treatment; endocrine abnormalities (e.g., hypothyroidism, unstable diabetes; history of central nervous system involvement (e.g., seizures, brain injury, neurological illnesses); any medical disorder believed by the investigator to cause depression; active suicidal risk or attempt during past year; pregnancy. Participants were recruited via newspaper advertisements that mentioned depression but not acupuncture, 2965 responded between January 1998 and March 2002

highlighting that acupuncture may be an acceptable intervention for pregnant women.

The majority of trials [26–28,35–40] have used cut-off points on the Hamilton Rating Scale for Depression (HRSD) for inclusion of patients with no discussion as to why a particular cut off point for entry, which ranged from a high of 25+ [38] down to 14+ [26], was selected. There was also an unexplained degree of variability in terms of depression diagnosis in the different studies: for example, Xuijuan et al. have investigated patients with manic and involutional depression, Luo et al. [35,36] have included patients with bipolar disorder and reactive depression, whilst the US and German studies have used tightly defined inclusion/exclusion criteria and have focused on major depressive disorder (MDD). With the exception of two studies [26,28] all trials have included both men and women. Allen et al. [28] focused on women only but did not provide reasons for piloting only on women and conducting the full-scale trial on both men and women. The fact that the pilot focused on women was described as a "study limitation" in the rationale for the full-scale trial. In terms of recruitment, all trials have (where reported) focused on hospital in-patients, with the exception of the US studies where recruitment was via newspaper and magazine advertisements. Generalizability of these patients to other contexts, such as primary care, where the majority of depression cases are managed and treated, may thus be problematic.

As highlighted in the introduction, previous illness history, particularly age of first onset and number of

Table 3 The intervention arms of depression studies of acupuncture

Trial (first author, date)	Acupuncture Intervention
1. Luo et al. 1985 (China)	Two acupuncture points (Du 20 and Yintang) were stimulated with electro acupuncture, six sessions a week for 5 weeks. Du 20 and Yintang were points used
2. Luo et al. 1988 (China)	Two acupuncture points (Du 20 and Yintang) were stimulated with electro acupuncture, six sessions a week for 6 weeks.
3. Xiujuan et al. 1994 (China)	Acupuncture points Du 24, 20, 14, 12, Ren 17, 14, GB 20, and P6 Additional points used depending on Chinese medical diagnosis Stagnation of Liver qi = St 23, Sp6, Liv3 Stagnation of Liver blood = LI4, Liv3, Sp10 Spleen and Heart deficiency = H7, P7, Sp6, St36 Spleen and Kidney yang deficiency = K3, Sp6, St36, Ren 4 Needles inserted bi-laterally and Du 24 and 20 stimulated with electro acupuncture (freq 80–100/second) Treatment administered for six consecutive days per week followed by a 1 day break for 6 weeks
4. Luo et al. 1998 (China)	Two acupuncture points (Du 20 and Yintang) were stimulated with electro- acupuncture (current 3–5 mA at a frequency of 2 Hz). Needles inserted obliquely 2–3 cm under the skin. Participants received six sessions a week for 6 weeks.
5. Allen et al. 1998 (USA)	Acupuncture based differential diagnosis according to Chinese medical concepts/theories. These were developed according to a treatment manual by the assessing acupuncturist (Schnyer). Patients were assessed by one acupuncturist whilst four others provided treatments. Practitioners were ostensibly blinded as to whether patients were receiving specific or nonspecific acupuncture, having rated their beliefs in treatment efficacy similarly for both types of intervention. Treatment involved twelve sessions over 8 weeks.
6. Roschke et al. 2000 (Germany)	Acupuncture, three times a week for 4 weeks. Delivered by 2 clinicians Standardized whole body treatment, three times per week over four consecutive weeks. Each session lasted 30 minutes, each patient being seen by the same practitioner. Points were U.B. 15, 17, 18, H7, P6, St40, Sp5, Sp6, and Lu1.
7. Han et al. 2002	All patients received daily electro acupuncture on Du 20 and Yintang for for 42 days) Patients given two additional points according to three different Chinese medical diagnoses
8. Manber et al. 2004	Acupuncture based differential diagnosis according to Chinese medical concepts/theories. These were developed according to a treatment manual developed by Schnyer. Patients were assessed monthly by one acupuncturist whilst others provided treatments. Treatment involved twelve sessions over 8 weeks. Responders received additional treatments
9. Allen et al. 2006	Acupuncture based differential diagnosis according to Chinese medical concepts/theories. These were developed according to a treatment manual developed by Schnyer. Patients were assessed by one acupuncturist whilst four others provided treatments which involved twelve sessions over 8 weeks

previous illness episodes, may be important prognostic indicators, and should ideally be assessed at baseline in order to ensure distribution is random across groups. This would be particularly important in small-scale studies. There was no report of illness history identified in the trial reports that were obtained for this review, or the Cochrane review.

The next topic considered by the review is the acupuncture care under investigation, and whether the trial is likely to be a fair and therefore useful evaluation of the therapy. Another important and related question is whether the care provided in the trial might be reflective of, and therefore generalizable to clinical practice.

The Acupuncture Intervention

Before discussing the nine trials in the review it is worth mentioning that there are considerable disparities in the way that acupuncture is routinely practiced in the West, where treatment is usually provided weekly/biweekly initially, continued over a few weeks or months with sessions that become more spaced out, and China, where a brief intensive daily course of therapy is more usual. Disparities in clinical practice between China and the West are reflected in the trials thus the interventions are substantially different in form. Table 3 summarizes the acupuncture intervention arms of the reviewed studies. Table 3 shows that Chinese studies (apart from Xiujuan [38]) have investigated standardized treatment protocols whereby participants were needled with the exact same combination of acupuncture points on each occasion. Han et al. [40] further divided patients allocated to acupuncture into three groups who were given different treatment (an additional two acupoints) depending on a Chinese medical diagnosis. The Roschke et al. [38] study used a standardized method whereby all patients received the same points.

The rationale for point selection in Luo et al. is basic; the points Bahui and Yintang were used because according to the authors these points are believed to be effective

for the relief of depression in Chinese medicine. In an earlier paper [37] Luo et al. provide a more detailed rationale including conducting preliminary research themselves on albino white rats by electro acupuncture stimulation of Bahui and Yintang and observing increases in 5HT and decreases in nor epinephrine. Further justification for the use of Bahui and Yintang acupoints was provided by observations of changes to a major metabolite of nor epinephrine in the urine of schizophrenic patients treated at these locations with electro acupuncture. Finally some quotes are taken from classic texts (*Suwen, Nan Jing (item 20), Jin Yue Quan Shu, Zhen Jui Da Cheng, Zhen Jiu Ji Cheng*) suggesting that in patients with "insanity" yin is greater than yang. This rationale, which appears to be an attempt to legitimize Chinese medicine and to replicate drug trial procedures, does not properly explain why Bahui and Yintang were chosen as the best points from the many point combinations that could potentially have been used.

Xiujuan et al. have provided a rationale for selection of points in the form of simple statements: for example: Neiguan, a crossing point of the Pericardium Channel of hand Jueyin and Chong Channel, benefits "the smoothing of the stomach and easing of mind." As with Luo et al. these authors present acupuncture points as if they have specific immutable properties that are equally true in every human at all times (or animal in the case of Luo et al. [37]) but this does not fit well with the *qi* paradigm outlined in the introduction. For example, according to acupuncture theory, the time of day, or season of the year may influence the healing potential of a particular acupoint, and some points may be more helpful for some individuals than others. Acupuncture is entirely context dependant and the context is not traditionally a specific disease or illness. Scheid [41] has highlighted how exponents of TCM have attempted to justify its underlying theories by aligning them with Western medical diseases. He proposes that in the development of TCM the concept of "principles" (*Li*) has been replaced by "theories" (*lilun*) [41]. These terms have entirely different meanings. The concept of "theory" has been imported to China from the West and through "theory" Chinese medicine and biomedicine have been conveniently, but artificially connected to each other, the same "theory" underpinning both.

Rationale for point selection was not reported by Roschke et al. [38].

All three US trials have used a manualiszed approach developed by Schnyer [42], which has been published in book format. In theory this looks to be a useful method. It is formulated around individualized treatments on the basis of diagnostic principles that involve looking, asking, listening and touch diagnosis and TCM pattern differentiation. The US studies have separated diagnosis from implementation of acupuncture needling in an innovative attempt to create a triple blind trial (blind patient, blind administering practitioner, and blind outcome assessor). The problem of discrepancies in practitioners' diagnoses, a characteristic feature of clinical practice, has been circumvented by using only one practitioner to make the diagnosis in Allen et al. [28]. However, the generalizability of the findings is compromised by the fact that the diagnostic skill of only one practitioner has been evaluated. Furthermore it is difficult to make judgements about quality of the intervention because it is not clear from the trial reports how many diagnosticians were involved or what their credentials were for Manber et al. and Allen et al. [26,28]. Manber et al. reported that patients were reassessed diagnostically once a month. It is not clear from the reports of either of the Allen et al. studies whether there was any diagnostic reassessment after the initial assessment was made. Thus patient's responses to therapy may not have been used to guide the treatment. Since implementing practitioners were discouraged from talking to their patients they may not have been able to gage whether there had been problems with the therapy. Additionally, acupuncturists invariably touch their patients, making assessments, particularly about meridians or responsiveness to treatment. It is therefore possible that the potential therapeutic integrity of the intervention may have been compromised by repeatedly using treatments that practitioners in normal clinical practice would have modified or altered because they were seen to be unhelpful. This limitation equally applies to any study where standardized protocol acupuncture is repeatedly administered. Acupuncture is an iterative process where the practitioner constantly reassesses the patient depending on their response to the treatment. In studies that use standardized protocols, an intervention that may be recognized as unhelpful or inadequate by the practitioner, and would normally be modified in routine clinical practice, would be continued for the duration of the trial. Furthermore, in the US studies, discrepancies between the information provided to administering practitioners and their blinding, and the reporting of study methods in relation to this, make it difficult, if not impossible, to interpret study findings. In the 1998 study it was reported that practitioners were blind as to the study's hypothesis, suggesting they may not have known that the trial was investigating depression. In the 2004 study authors reported that practitioners were informed that the study was investigating points that were directly and indirectly relevant for depression whilst in the 2006 trial practitioners were told that the study was investigating various forms of treatment for depression. Researchers concluded that blinding of administering practitioners

Table 4 Comparator arm interventions used in acupuncture trials of depression

Trial (first author, date)	Comparator Arm(s)
1. Luo 1985 (China)	Amitriptyline medication (25 mg, 3× per day for 1 week then increased to an average dose of 142 mg)
2. Luo 1988 (China)	Amitriptyline medication for 6 weeks. Initial dose for week one: 25 – 50 mg, increased to 150 mg per day. From week 2 the dose was modified depending on side effects.)
3. Xiujuan 1994 (China)	Amitriptyline 25 mg on the first day, dose increased by 25–50 mg each day up to 150 mg. In the second week the dose was adjusted according to response and side effects but ranged between 150 and 300 mg per day
4. Luo 1998 (China)	Phase 1: Amitriptyline treatment group received average dose of 175 mg per day, combination group received 165 mg per day. Phase 2:Amitriptyline given at average dose of 161 mg per day.
5. Allen 1998 (USA)	Two comparator arm interventions: Nonspecific acupuncture using acupuncture points but for another diagnosis e.g., back pain, and wait list control. The nonspecific acupuncture and wait list control groups received specific acupuncture after 8 weeks. Administering acupuncture practitioners were blind as to study's hypothesis.
6. Roschke 2000 (Germany)	Three groups Mianserin (90–120 mg per day) Mianserin (90–120 mg per day) plus verum acupuncture Mianserin (90–120 mg per day) plus placebo acupuncture Placebo acupuncture patients received treatment at nonacupuncture loci adjacent to the selected acupoints and the skin was only superficially pricked.
7. Han 2002	Maprotyline medication, daily dose ranging from 75–250 mg, for 6 weeks
8. Manber et al. 2004	Two comparator arm interventions Nonspecific acupuncture-participants needled at nonrelevant points Massage, as a time and attention control, exploring the variable of physical contact. Administering acupuncturists were told the study was investigating points that were directly and indirectly relevant for depression.
9. Allen et al. 2006	Two comparator arm interventions: Nonspecific acupuncture using acupuncture points, and wait list control. The nonspecific acupuncture and wait list control groups received specific acupuncture after 8 weeks. Administering acupuncture practitioners were told the study was investigating various forms of treatment for depression.

may not have been entirely successful in this final 2006 trial. It is difficult to know what to make of these studies because in clinical practice practitioners know why their patients have come to them for help and what the purpose of their treatment is meant to be. Communication, which practitioners were requested to keep to a minimum in US studies, is an important aspect of acupuncture care, particularly in cases of depression where patients may require a considerable amount of support from the practitioner. Authors of the three US studies acknowledge the ecological validity of their research is low and their findings cannot be generalized to standard acupuncture care.

This critique has highlighted potential limitations resulting from trialists' attempts to standardize care in order to achieve methodological rigor. The advantage of standardization must be offset against the disadvantage that such treatments, when obviously inadequate or inappropriate, cannot be modified, as would normally occur in routine clinical practice. Finally, in the light of the discussion in the introduction about the many diverse traditions of acupuncture that are practiced in the world today [20], it is worthwhile highlighting to the reader that all the research attention in the field of depression conducted so far has been focused on TCM methods, and techniques developed from TCM theories. Other traditionalist methods have yet to be formally evaluated.

The next section of the review will examine the comparator arms used by trialists, outlining what these studies can be expected to reveal about acupuncture's potential effectiveness.

Comparator Arm Interventions

Apart from the three USA studies, all trials have used a comparator arm intervention of antidepressants (see Table 4). Cochrane reviewers have raised concerns that doses were therapeutically ineffective in the Luo trials and the timing of assessment was premature [43]. In any experimental evaluation it is important that the therapy under investigation is not privileged by providing a therapeutically inadequate comparison. Roschke et al. [39] have used acupuncture as an adjunctive treatment with Mianserin. A placebo acupuncture intervention was also used, where patients were superficially needled at nonacupuncture loci adjacent to "real" acupoints. Pilot work was not undertaken to make sure the full-scale trial was adequately powered to compare sham

Table 5 Outcome measures used in trials of acupuncture for depression

Trial (first author, date)	Outcome measures (measures used and when measured)
1. Luo 1985 (China)	Hamilton Rating Scale for Depression, Clinical Global Impression Chart, Antidepressant side effect rating scale. Participants interviewed by 2 psychiatrists ant beginning and end of trial (5 weeks), and at weekly intervals during the trial
2. Luo 1988 (China)	Hamilton Rating Scale for Depression, Clinical Global Impression Chart, Asberg Rating Scale for Side-effects. Grading System for assessment of therapeutic effects commonly used in China (cured, markedly improved, improved, failed/deteriorated). A proportion of participants had urine and plasma tests and EEG tests
3. Xiujuan 1994 (China)	Hamilton Rating Scale for Depression used once a week for 6 weeks. EEG measures used for acupuncture subjects only at baseline, after 1st, 18th and final needling
4. Luo 1998 (China)	Hamilton Rating Scale for Depression, Clinical Global Impression Chart, Asberg Rating Scale for Side-effects. Grading System for assessment of therapeutic effects commonly used in China (cured, markedly improved, improved, failed/deteriorated). A proportion of participants had urine and plasma tests and EEG tests in 2nd phase of research
5. Allen 1998 (USA)	Hamilton Rating scale for Depression and Becks Depression Inventory at baseline, 8 and 16 weeks.
6. Roschke 2000 (Germany)	The Global Assessment Scale (Endicott et al. 1976), Bech-Rafaelsen Melancholia Scale, Clinical Global Impressions Scale (National Institute Mental Health 1976), self report of improvement (von Zerssen et al., 1974), mean dosage of medication. Rated by psychiatrists blind to allocation twice a week for 8 weeks
7. Han 2002	Hamilton Rating Scale for Depression, self rating scale for depression (unspecified), self rating scale for anxiety (unspecified) Clinical Global Impression Scale, Asberg Rating Scale for Side-effects. Baseline, 14, 28, and 42 days after treatment commenced
8. Manber et al. 2004	Main outcome measure was response status at end of acute treatment phase. Response was defined jointly by a) failure to meet full criteria for MDD; b) at east 50% reduction from baseline of HRSD score and c)HRSD score of 14 or less. HRSD administered at baseline, midpoint, end of acute treatment (8 weeks) and 10 weeks postpartum by interviewers masked to treatment assignment MDE portion of SCID administered at same time as HRSD by same interviewers Becks Depression Inventory completed weekly during acute treatment phase. Providers and patient's expectations of efficacy of treatments assessed at first and third treatments
9. Allen et al. 2006	HRSD was main outcome measure, taken at baseline and 4 weekly intervals throughout study period. Administered by assessors blind to allocation. Becks Depression Inventory (BDI) used weekly for 16-week study duration.

and "real" acupuncture. Knowing the effect size of an acupuncture intervention in relation to a sham comparator is important to avoid a false negative result [44].

All three US studies, the first of which [28] was described as a pilot study, have employed what is described as a nonspecific acupuncture comparator arm, where acupuncture points that were not thought to be therapeutically appropriate were needled. Although the Allen et al. [28] study mentioned the use of points for back pain instead of depression as an exemplar, the rationale for nonspecific acupuncture point selection has been inadequately reported in other studies. It is not clear what points were used, why they were chosen, and whether they varied from patient to patient or between treatment sessions. Acupoints that are contraindicated during pregnancy according to TCM theory were avoided in the Manber trial but there is a paucity of knowledge in this area because systematically conducted safety research has not, to authors' knowledge, been conducted that has specifically focused on pregnancy treatment with acupuncture. Subjecting depressed pregnant women to a theoretically inappropriate treatment, which has apparently achieved worse outcomes than no intervention at all in piloting work, is possibly ill advised. Manber et al. [26] employed massage as a control intervention for attention, physical contact, relaxation, and respite from daily stress. Verbal contact was kept to a minimum in all trial arms/consultations. But in clinical practice acupuncture practitioners and massage therapists do not have to keep verbal interactions to a minimum, and use verbal interaction to gain and maintain rapport with their patients.

The final section of the review will be on outcomes. What instruments were used and when measurements were taken, an overview being shown in Table 5.

Outcomes

As was discussed in the introduction section, researchers have emphasized that an intervention for depression should be evaluated according to its capacity to offer protective benefits, as well as to achieve a more rapid or

improved remission from a specific episode [16,45]. Because of the chronic episodic nature of depression, the potential economic advantages of an intervention are likely to become more apparent in the long term if the intervention can prevent relapse (in addition to achieving remission in the short term). With the exception of Manber et al. [26], no acupuncture study has explored outcomes beyond a few days or weeks after treatment ceased.

In terms of outcome measures Table 5 shows that the Hamilton Rating Scale for Depression has been used as the primary outcome measure in all studies apart from Roschke et al. The Luo et al. [35,36] have reported using biochemical and EEG tests to assess a proportion of patients. How patients were selected for these tests was not reported.

Broader outcomes of acupuncture, such as quality of life, and economic benefits of the intervention have not been explored. In clinical practice it would be unusual for acupuncture to be focused solely on a specific medical condition, and a wide range of health benefits, are normally be expected from treatment. For example it would be expected that a female patient with depression who also suffered with irregular or painful menstruation would find that her menstrual cycle became more regular and less painful. From the patient's perspective, the potential benefits of acupuncture for individuals with chronic illnesses have been investigated by Paterson and Britten with changes falling into three main categories: changes to health status including improvements to a specific condition and improved strength and energy; changes to social and personal identity; and changes to medication use [46]. Ideally, a range of measures are likely to be required to capture the potential benefits of acupuncture [47].

Discussion

Summary of Findings

The study population, intervention, comparison(s), and outcomes of nine trials have been evaluated with our review highlighting a number of study limitations, quite apart from methodological considerations that prevent bias and confer reliability. One of the main concerns is that in attempting to evaluate replicable treatments, interventions have been standardized, and protocol acupuncture, where individuals receive the same or similar points without modification during the course of the trial have been investigated. Acupuncture care that is modified in line with patient's responses, as would occur in routine clinical practice, has not been evaluated. Secondly, in some trials where a flexible protocol was evaluated, notably where the quality rating was highest according to Cochrane reviewers, treatment was administered by practitioners who were discouraged from talking to their patients and did not appear to know, according to trial reports, exactly why they were treating them. Concerns were raised about the use of therapeutically inappropriate treatment as a safe or useful comparison. It is possible, given individual variations in responsiveness to acupuncture, there may be a degree of unpredictability about this intervention that makes it unsuitable for comparative purposes. Further problems in the evidence base were identified with outcomes focused solely on depression and on the immediate or short-term impact of interventions. Protective benefits of acupuncture care were not investigated, yet, as was highlighted in the introduction, this is where the economic benefits of a depression intervention may be most apparent. Additional research costs associated with longer term treatment and follow up can and should be offset against the potential reduction in the cost to society, and suffering caused to individuals, of repeated illness episodes.

Discussion of Findings in the Light of Other Research

It is evident that the evaluations of acupuncture for depression in this review can be said to have poor "model validity." The concept of model validity, proposed by Verhoef et al. [48], suggests that any evaluative research should "adequately address an intervention's unique healing theory and therapeutic context" [48]. Authors suggest that model validity is crucial for whole systems research and unconventional therapies such as acupuncture where philosophical assumptions may differ from those of biomedicine. Four criteria have been proposed for assessment of model validity by Verhoef et al. These are: *representativeness*—whether the intervention is consistent with current practice, likely to produce and effect and clearly described; *equipoise and credibility*—to assess equipoise on the part of patients and practitioners with respect to the intervention under investigation and its credibility in relation to expected treatment effects; *model congruity*—to assess whether the diagnosis, intervention and outcomes fit the system under investigation, and finally *context*—to assess patient and practitioner confidence in the intervention and whether the intervention was sensitive to the culture, family and meaning that the patient ascribes to an intervention. Clearly, the nine trials in this overview fall short on these criteria. However, whilst it is difficult to justify the expenditure of scant research resources on evaluations of potentially suboptimal treatments where "model validity" is a poor fit, the research into acupuncture for depression has suggested that acupuncture, when implemented according to

traditional Chinese medical theories, holds promise, and specific techniques may be as effective in the short term as antidepressants. The problem is that the potential value of TCM acupuncture as a depression intervention will remain unknown if evaluations do not allow practitioners to implement their therapy as they would in everyday clinical practice. However, it is not just acupuncture trials of depression that have been criticized for providing inadequate information on which to base decisions about provision of care. RCTs of antidepressants have also come in for criticism because medications produce different results in tightly controlled trials than they do under normal practice conditions [49]. Reviewers of mental health research have called for more pragmatic RCTs "comparing routine packages of care about which there is genuine uncertainty and measuring outcomes of interest to clinicians and patients" [50].

Whilst the purpose of this review was not primarily to discuss the merits of using specific acupuncture protocols or sham or placebo acupuncture for comparison, it is worth highlighting that research has suggested that practitioners working in trials where "protocol" acupuncture is investigated (the most common form of study design for acupuncture depression trials) find it a frustrating experience, as they are not able to treat holistically, or make their own diagnostic judgments about what is appropriate [25]. Research has also revealed how it can be extremely uncomfortable for practitioners to participate in explanatory trials of acupuncture because of being aware that the patient is receiving what is thought to be inadequate treatment and needing to conceal this fact from them in the context of trying to maintain a close therapeutic relationship during a number of therapy sessions [51]. Evidently nonspecific interventions involving the insertion of needles are more potent than placebo pills [52]. Although acupuncture's theories remain unproven it is questionable whether it is in the best interests of patient safety for these theories to be completely disregarded, and needles inserted into therapeutically inappropriate locations, particularly in the case of patients who are pregnant, and obviously have a high risk of suicide as a result of their illness. Such interventions are not likely to be physiologically "inert," and results obtained from these studies may be confusing and difficult to interpret. If acupuncture (hypothetically) has the potential to heal does it not (hypothetically) have the potential to harm if inappropriate treatment is given?

Meaning of Findings and Implications for Future Research

One way forward for future research is to evaluate acupuncture as the complex intervention that it is. In certain respects acupuncture is similar to counseling or psychological therapies in that the therapeutic relationship, which develops over time, cannot be separated from the intervention. This relationship is built on patients feeling comfortable with their practitioner and the explanations of illness specific to Chinese medicine their practitioner offers. Traditional acupuncture may, as outlined in the introduction, involve a process of bodily "reducation" or learning, and in this respect it is not dissimilar to other health educational interventions, except for the Chinese theories that underpin it. Guidelines for complex intervention evaluations such as psychological therapies and health education interventions have been put forward by the Medical Research Council [53], and may be of interest to trialists thinking of how best to evaluate traditional acupuncture. One of the most challenging aspects of conducting complex intervention research is undoubtedly the standardization of the intervention, which should according to researchers, be done on the basis of function rather than form [54]. It is important to point out that the lack of standardization does not actually compromise the internal validity of a study [55], it just makes it "noisy" and more difficult to interpret the results. But researchers will need to meet the challenge if they are to evaluate acupuncture in such a way as to be reflective of clinical practice. Trials of specific acupuncture protocols are nevertheless of interest as they, when methodologically sound, offer convincing evidence of the potential for specific acupuncture techniques to be beneficial, and when brain scan research demonstrates biological changes resulting from interventions (as suggested by the most recent systematic review [33] this is likely to give acupuncture more credibility as a potentially useful intervention for depression.

Conclusion

Depression is a huge and increasing problem that costs society a great deal of money, and impacts on many individuals, directly and indirectly. Considerable uncertainty remains about whether acupuncture can be an effective intervention for depression—despite a main reason for help seeking with acupuncture being described as "mood care." It has been argued that the current evidence base has not adequately addressed these uncertainties. Research is required that can actually inform patients and service commissioners as to whether the treatment they might, in practice, receive from a therapist is potentially likely to be of any clinical value, and whether it may be harmful. Trials of acupuncture are needed that investigate the effectiveness of acupuncture, in such a way that the potential therapeutic integrity of the intervention is

not compromised by the evaluation method, whilst maintaining the rigor of the evaluation. A recommendation of this review is that research is conducted in the field of depression with better model validity, and that can be of more practical value for informing decisions about provision of care. Unless acupuncture is evaluated using methods that do not compromise its potential therapeutic integrity by limiting its evaluation to specific acupoints applied outside the context of a proper Chinese medical diagnosis, uncertainty will remain about whether it can truly be effective or not for treating depression.

Acknowledgments

We would like to thank Trevor Sheldon, Hugh MacPherson and Simon Gilbody. Thanks also to Karen Smith who provided advice on searching the literature and to Wenqing Li who translated two Chinese articles into English. Sylvia Schroer was the recipient of a personal award granted by the UK National Coordinating Centre for Research Capacity Development.

Conflict of Interest

The authors have no conflicts of interest.

References

1. WHO. World Health Report 2001, Mental Health: New Understanding, New Hope. Geneva, Switzerland: World Health Organisation, 2001.
2. Thomas C, Morris S. Cost of depression among adults in England in 2000. *Br J Psychiatry* 2003;**183**:514–519.
3. Chisholm D. The economic consequences of depression. In: Dawson A, Tylee A, editors. *Depression: Social and economic timebomb: Strategies for quality care*. London: BMJ Books on behalf of The World Health Organisation, 2001.
4. Goldberg D, Privett M, Ustun B, Simon G, Linden M. The effects of detection and treatment on the outcome of major depression in primary care: A naturalistic study in 15 cities. *Br J Gen Practice* 1998;**48**:1840–1844.
5. Fava M. Diagnosis and definition of treatment resistant depression. *J Biol Psychiatry* 2003;**53**:649–659.
6. NICE. *Depression: Management of depression in primary and secondary care*. London: NCCMH, 2004.
7. Fisher P, Van Haselen R, Hardy K, Berkovitz S, McCarney R. Effectiveness Gaps: A new concept for evaluating health service research needs applied to complementary and alternative medicine. *J Alternative Compl Med* 2004;**19**:627–632.
8. Hornsby R. Depression—A patients view of the future. In: DAaT A, editor. *Depression: Social and economic timebomb*. World Health Organisation: BMJ Books, 2001.
9. Barnes LL. The psychologizing of Chinese healing practices in the United States. *Cult Med Psychiatry* 1998;**22**:413–443.
10. Cassidy C. Chinese Medicine Users in the United States. Part 2: Preferred aspects of care. *J Alternative Comp Med* 1998;**4**:189–202.
11. Judd LL. The clinical course of unipolar major depressive disorders. *Arch Gen Psychiatry* 1997;**54**:989–991.
12. Segal Z, Pearson J, Thase M. Challenges in preventing relapse in major depression. Report of a National Institute of Mental Health Workshop on state of relapse prevention in major depression. *J Affect Disord* 2003;**77**:97–108.
13. Keller MB, Lavori PW, Lewis CE, Klerman GL. Predictors of relapse in major depressive disorder. *J Am Med Assoc* 1983;**250**:3299–3304.
14. Giles DE, Jarrett RB, Biggs MM, Guzick DS, Rush AJ. Clinical predictors of recurrence in depression. *Am J Psychiatry* 1989;**146**:764–767.
15. Katon W, Schulberg H. Epidemiology of depression in primary care. *Gen Hosp Psychiatry* 1992;**14**:237–247.
16. Goldberg D, Goodyer I. *The origins and course of common mental disorders*. New York and London: Routledge, 2005.
17. Unschuld P. Interview with Paul Unschuld. *European J Orient Med* 1993;**1**:7–11.
18. Hsu E. *The transmission of Chinese medicine*. Cambridge: Cambridge University Press, 1999.
19. Scheid V. *Chinese medicine in contemporary China: Plurality and synthesis*. Durham, North Carolina: Duke University Press, 2002.
20. Birch S, Felt R. *Understanding acupuncture*. Edinburgh: Churchill Livingston, 1999.
21. Schnyer R, Birch S, MacPherson H. Acupuncture practice as the basis for evaluation. In: MacPherson H, Hammerschlag R, Lewith G, Schnyer R, editors. *Acupuncture research strategies for developing and evidence base*. Edinburgh: Churchill Livingstone, 2008.
22. Scheid V, Bensky D. Medicine is signification: Moving towards healing power in the Chinese medical tradition. *Eur J Orient Med* 1998;**2**:32–41.
23. Hicks A. Intention- The art of 'doing nothing'. *Chin Whispers*. 2007; Winter Issue.
24. Bensoussan A. *The vital meridian: A modern exploration of acupuncture*. Melbourne: Churchill Livingston, 1991.
25. Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 2005;**330**:1202–1205.
26. Manber R, Schnyer RN, Allen JJB, Rush AJ, Blasey CM. Acupuncture: A promising treatment for depression during pregnancy. *J Affect Disord* 2004;**83**:89–95.
27. Allen J, Schnyer R, Chambers A, Hitt S, Moreno F, Manber R. Acupuncture for depression: A randomized controlled trial. *J Clin Psychiatry* 2006;**67**:1665–1673.
28. Allen J, Schnyer R, Hitt S. The efficacy of acupuncture in the treatment of major depression in women. *Psychol Sci* 1998;**9**:397–401.

29. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence based medicine: How to practice and teach EBM*. 2nd ed. Edinburgh: Churchill Livingston, 2000.
30. Mukaino Y, Park J, White A, Ernst E. Effectiveness of acupuncture for depression: A systematic review of randomised controlled trials. *Acupunct Med* 2005;**23**:70–76.
31. Leo R, Ligot J. A systematic review of randomized controlled trials of acupuncture in the treatment of depression. *J Affect Disord* 2007;**97**:13–22.
32. Wang H, Qi H, Wang B-S, Cui Y-Y, Zhu L, Rong Z-X, Chen HZ. Is acupuncture beneficial in depression: A metaanalysis of 8 randomized controlled trials. *J Affect Disord* 2008;**111**:125–134.
33. Zhang Z-J, Chen H-Y, Yip K, Ng R, Wong V. The effectiveness and safety of acupuncture therapy in depressive disorders: Systematic review and meta-analysis. *J Affect Disord* 2009. http://dx.doi.org/10.1016/j.jad.2009.07.005.
34. Sackett DL, Cook RJ. Understanding clinical trials.[comment]. *BMJ* 1994;**309**:755–756.
35. Luo HC, Shen Y, Jia Y, Zhou D. Clinical study of electro-acupuncture on 133 patients with depression in comparison with tricyclic amitriptyline. *Chin J Modern Develop Tradition Med* 1988;**8**:77–80.
36. Luo HC, Meng FQ, Jia YK, Zhao XY. Clinical research on the therapeutic effect of the electro-acupuncture treatment in patients with depression. *Psychiatry Clin Neurosci* 1998;**52**:S338–S340.
37. Luo HC, Yunkui J, Zhan L. Electro-acupuncture vs amitriptyline in the treatment of depressive states. *J Tradition Chin Med* 1985;**5**:3–8.
38. Xiujuan Y., Xiang L, Hecun L, Jia Y. Clinical observation on needling extra-channel points in treating mental depression. *J Tradition Chin Med* 1994;**14**:14–18.
39. Roschke J, Wolf C, Muller MJ, Wagner P, Mann K, Grozinger M, Bech S. The benefit from whole body acupuncture in major depression. *J Affect Disord* 2000;**57**:73–81.
40. Han C, Li X, Luo H. Comparative study of electro-acupuncture and maprolitine in treating depression. *Comb Chin/Western J* 2002;**22**: 512–514.
41. Scheid V. Not very traditional, nor exactly Chinese, so what kind of medicine is it? TCM's discourse on menopause and its implications for practice, teaching and research. *J Chin Med* 2006;**82**:5–20.
42. Schnyer R, Allen J. *Acupuncture in the treatment of depression*. Edinburgh: Churchill Livingston, 2001.
43. Smith CA, Hay PPJ. Acupuncture for depression. *Cochrane Database System Rev* 2004. doi:10.1002/14651858.
44. Birch S. Controlling for non-specific effects of acupuncture in clinical trials. *Clin Acupunct Orient Med* 2003;**4**:59–70.
45. Roth A, Fonagy P. *What works for whom*. New York, London: Guildford Press; 2005.
46. Paterson C, Britten N. Acupuncture for people with chronic illness: Combining qualitative and quantitative outcome assessment. *J Altern Complement Med* 2003;**9**:671–681.
47. Paterson C. Measuring changes in self-concept: A qualitative evaluation of outcome questionnaires in people having acupuncture for their chronic health problems. *BMC Complement Alternat Med* 2006. doi:10.1186/1472-6882-6-7.
48. Verhoef MJ, Lewith G, Ritenbaugh C, Boon H, Fleishman S, Leis A. Complementary and alternative medicine whole systems research: Beyond identification of inadequacies of the RCT. *Complement Therap Med* 2005;**13**:206–212.
49. Simon G, Wagner E, Vonkorff M. Cost-effectiveness comparisons using real-world randomized trials—The case of new antidepressant drugs. *J Clin Epidemiol* 1995;**48**:363–373.
50. Adams C. What have we learned from 50 years of randomised trials for people with schizophrenia. In: Duley L, Farrell B, editors. *Clinical trials*. London: BMJ Books, 2002.
51. McManus C, Kaptchuk T, Schnyer R, Goldman R, Kerr CE, Nguyen L, Stason WB. Experiences of acupuncturists in a placebo-controlled, randomized clinical trial. *J Alternat Complement Med* 2007;**13**:533–537.
52. Kaptchuk TJ, Stason WB, Davis RB, et al. Sham device versus inert pill: Randomised controlled trial of two placebo treatments. *Br Med J* 2006;**332**:391–394.
53. Medical Research Council. A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health. MRC 2000.
54. Hawe P, Shiell A, Riley T. Complex interventions: How "out of control" can a randomised controlled trial be?. *BMJ* 2004;**328**:1561–1563.
55. Elwood M. *Critical appraisal of epidemiological studies and clinical trials*. 3rd ed. Oxford: Oxford University Press, 2007.