

Acupuncture is ineffective for chronic low back pain? A critical analysis and rethinking

Xuqiang Wei¹, Baoyan Liu (✉)^{1,2}

¹College of Acupuncture and Orthopedic, Hubei University of Chinese Medicine, Wuhan 430065, China; ²Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China

© Higher Education Press 2021

Abstract Acupuncture is a promising treatment for relieving pain and improving lower back function in clinical practice. However, evidence from randomized clinical trials (RCTs) remains controversial. Most RCTs conclude that acupuncture procedures for chronic low back pain (CLBP) had no significant difference in efficacy and belonged to placebo. We carefully reviewed and analyzed the methodology and implementation of sham acupuncture in RCTs. Controversial evidence of acupuncture for CLBP is only a microcosm of the evaluation methodological limitation of acupuncture. Inappropriate selection of sham acupuncture controls, rigorous RCT research models, and incorrect interpretation of results may contribute to negative evidence. Evaluating and disregarding the holistic efficacy of acupuncture with an explanatory RCT model based on evaluation drugs may be unwise. Moreover, sham acupuncture is often proven to be non-inert, unreasonable, and with low fidelity. Pitfalls of the explanatory RCT model and sham acupuncture design should be avoided. Establishing a new evaluation system that is in line with the clinical characteristics of acupuncture and obtaining high-quality evidence are difficult but promising tasks.

Keywords chronic low back pain; sham acupuncture; acupuncture; methodology; therapeutic evaluation

Backgrounds

With the acceleration of global aging, the increase in obesity, the sedentary lifestyle of people, chronic low back pain (CLBP) has become a global disease with a high incidence [1]. According to the Global Burden of Disease Study (GBD) 2017, low back pain is the second leading cause of disability-adjusted life years in non-communicable diseases [2]. The increased disability rate of CLBP has weakened labor capacity, reduced the quality of life, and increased the psychological, social, and medical burden. Unfortunately, 85%–95% CLBP cases are not attributed to specific pathoanatomical origin or recognizable pathology patterns [3]. Such lack of specific treatment target leads to the creation of an inefficient biomedical model of CLBP management. Therefore, an effective and safe treatment is especially important for overcoming low back pain and disability associated with a chronic condition [4].

Over the past three decades, the main treatment recommendations in the national clinical practice guide-

lines for CLBP have changed. Opioids and analgesics are currently discouraged, because evidence of efficacy is unclear, and these substances can be harmful. Intervention procedures and surgery play a very limited role in CLBP [5]. Unexpectedly, acupuncture and other types of non-pharmacological therapy have been receiving an increasing amount of attention as treatment options [5–7]. Acupuncture is a typical non-drug therapy that is used to address the opioid crisis. It is especially being considered for CLBP relief. It originated from ancient China and is accepted to be an effective, safe, and reliable intervention for CLBP in many countries [8,9]. A cross-sectional study has shown that CLBP is the most commonly treated condition with acupuncture in the United States [10].

Alleviating pain and improving lower back function effectively without serious harm are the effective indicators for CLBP treatment. Acupuncture has potential as an effective intervention. Recently, an individual level meta-analysis of acupuncture for chronic pain demonstrated the effectivity of acupuncture, and its analgesic effect remains stable and durable over time [11]. Data analysis and synthesis from 20 827 patients in 39 trials showed that the effects of acupuncture cannot be interpreted simply as placebo. According to analysis, the analgesic effect of

acupuncture lasted for more than 1 year among 90% of the patients compared to the control groups [11]. Another authoritative systematic review (SR) released by the Agency for Healthcare Research and Quality (AHRQ) highlighted that acupuncture is effective for CLBP with minimal side effects [12]. Moreover, a randomized sham-controlled trial suggested that acupuncture can reduce botheration and pain intensity significantly in patients with CLBP [13]. The evidence on the effectivity of acupuncture for CLBP seems convincing.

Nevertheless, acupuncture has been interpreted as a powerful “placebo” in some RCTs and a meta-analysis [14–16]. In 2016, the National Institute of Health and Clinical Excellence (NICE) published a guideline for low back pain diagnosis and treatment. Acupuncture is not recommended for CLBP on the basis of insufficient evidence [15]. This finding provoked debate on the effectiveness and safety of acupuncture as a treatment for low back pain [17,18]. Many researchers argued that acupuncture is ineffective compared with sham acupuncture. Although this conclusion may be attributed to methodological heterogeneity, acupuncture has not been objectively evaluated. The conclusion that acupuncture has no specific effect or is only a “placebo” may be arbitrary and not credible. Acupuncture has been used in China to treat low back pain more than 2000 years. Many patients with CLBP have been treated with acupuncture. Nevertheless, in the era of evidence-based medicine, the evaluation of the effectiveness of acupuncture as a therapeutic intervention for CLBP is controversial [19]. Compared with the positive image of acupuncture in the minds of acupuncturists, evidence on the efficacy of acupuncture is surprisingly negative. What caused the current inconsistent situation? Aside from the debate, should the acupuncture clinical trial’s scientific, objective, and comprehensive nature be scrutinized as well? Most of the evidence originated from RCTs, in which the control groups are different kinds of sham acupuncture. Two possible explanations for the abovementioned findings are evident; either acupuncture was indeed an effective placebo or the negative result was wrongly interpreted [20]. False-negative results can occur for a variety of reasons [21], including contingency, insufficient sample size, improper outcome selection, inadequate intervention, and inappropriate control procedures. Nevertheless, the methodological flaws, unreasonable comparison, and study population deserve a second look.

Methodological challenge in acupuncture research

Traps in RCT models

The main advantages of RCT are the reduction of selection

bias and the control of confounding factors through random selection and blind method. Thus, RCTs have become the “gold standard” for testing intervention effects. However, acupuncture studies that comply with the classical “gold standard” of randomized, double-blind, placebo, and sham-controlled designs may encounter some methodological pitfalls [22]. Frequent inactive effects compared with putative “inert” placebos are inconsistent with actual clinical practice of acupuncture for CLBP. As a complex intervention, acupuncture is difficult to incorporate into the curative effect evaluation system of RCTs. Therefore, it is necessary to reflect on the rationality of the design of RCTs in a clinical research involving acupuncture.

As an individualized, dynamic, and holistic treatment style, syndrome differentiation is the core basis of acupuncture. Throughout the acupuncture treatment process, the treatment scheme is continuously adjusted according to the changes in the individual’s condition. The limitation of clinical research methodology has been an obstacle in the development of acupuncture efficacy evaluation, specifically in the following aspects.

First, timely patient feedback, adequate doctor-patient interaction, and flexible acupuncture treatment strategies are important factors in the entire personalized acupuncture process. An RCT research model highlights the effect of intervention on the homogenized group, which contradicts the concept of individualized acupuncture treatment. Second, acupuncture is a complex overall intervention. In addition to acupoint selection, the effect is also affected by many potential factors, such as doctor’s skills, mutual trust between doctors and patients, patient expectations, treatment environment, needle specifications, and acupuncture operation [23]. Evaluating the specific efficacy of acupuncture only by controlling its parameters by randomization and blinding methods is not enough. Third, it is almost universally accepted that that acupuncture treatments directly operated by doctors cannot achieve high reliability and validity under the blind method. Sham acupuncture is obviously contrary to the strict blindness requirements of classic RCTs. Therefore, acupuncture may not be suitable for the RCT research paradigm; it is more suitable for drug and biomedically oriented interventions [24].

To design a placebo or sham-control RCT, acupuncture must be divided into two parts, namely, characteristic (specific) and accidental (placebo, non-specific). However, classification characteristics and accidental factors in acupuncture have been revealed as meaningless. Accidental factors in drug trials may be a part of non-pharmacological interventions [25]. Evidence and conclusions from RCTs may not be in line with the essence of acupuncture and clinical practice. Accordingly, strictly following the RCT scale may lead to the loss of the characteristics of acupuncture and hinder its development.

In addition, according to the criteria of RCTs, studying the effects of acupuncture on the population level may be an impractical or unreasonable solution.

Traps in the study population

CLBP is a common symptom that affects nearly all age groups; the cause is non-specific nociceptive [1]. RCTs of CLBP should cover the entire population. However, RCTs are often conducted among specific populations with strict inclusion and exclusion criteria; the specific settings are out of the real clinical practice environments [26]. CLBP diagnosis is based on symptoms or complaints, and many specific or non-specific factors may be related to back pain. Different RCTs seem to investigate the same CLBP diagnosis, but the condition may involve different diseases. Different acupuncture schemes or fake acupuncture designs have different effects on various randomized controlled trials. Moreover, the population in RCTs and the acupuncture operators may be inconsistent with the real clinical settings. The acupuncturist's practice level, clinical experience, and acupuncture program selection can directly affect the results of the study. Nearly all acupuncturists may realize that variation in effectiveness is inevitable even with the same points and the same techniques. Therefore, trying to compare the effect of individual acupuncture treatments on CLBP at the population level in RCTs is a big pitfall.

Traps in efficacy evaluation of acupuncture

Drug development involves the cautious introduction of a new substance into the human body. Progressive assessment of the patient's disease is performed to evaluate the drug's safety, efficacy evidence, and appropriate doses for future evaluation. In this model, clinical efficacy evaluation usually relies on prospective and confirmatory trials of humans based on preclinical animal studies. To obtain reliable and valid evidence for human body application, the strict control of various factors is needed to create an ideal clinical trial environment. However, acupuncture originates from clinical practice. In the process of solving clinical problems, experience is continuously accumulated, and solutions are optimized. Acupuncture has gradually evolved into group technology and theory through the iteration of individual experience.

The efficacy evaluation of acupuncture differs from that of new drugs. New drug evaluation tends to be confirmatory and explanatory, whereas acupuncture evaluation is inclined to be optimal and pragmatic. Explanatory trials confirm physiological or clinical hypotheses that indicate the specificity of a new drug or treatment. Pragmatic trials inform clinical or policy decisions by providing the evidence of the application of intervention to actual clinical practice. Unfortunately, due to the limita-

tions of the explanatory RCT thinking model, a placebo-controlled RCT design is still unreasonably used in clinical trials involving the use of acupuncture for CLBP, even if clinically validated evidence is abundant. Many RCTs on the use of acupuncture for CLBP have neither optimized the clinical practice of acupuncture nor demonstrated the superiority of the treatment. Under the influence of sham acupuncture and interpretation bias, researchers assert that acupuncture does not show a specific therapeutic effect. In the RCTs, the specific effects of acupuncture are not always distinguishable from placebo effects. False-negative results may be generated when we adopt a placebo or sham-controlled trial design to detect the whole characteristic effect [25].

For complex interventions like acupuncture, the efficacy evaluation should be systemic, and the traditional double-blind placebo trial may not be appropriate. The diversity of treatment compositions, social attributes, and individual differences of acupuncture treatment subjects should be fully considered. To evaluate acupuncture objectively and comprehensively, new methods with strong applicability and that do not deviate from strict design principles need to be developed.

Some scholars believe that the controversy over the evaluation of acupuncture efficacy may be due to the inevitable heterogeneity of acupuncture. Ideally, the heterogeneity may be conquered through a "placebo needles" design, which is supposed to control the placebo effects of acupuncture adequately [27]. However, even by using sham devices, many studies fail to show the difference of effects among various acupuncture groups. This conclusion has evoked considerations regarding the design of "inert" placebo (or sham) acupuncture and clinical research methods.

Challenges in acupuncture comparison

Methodology difficulty of placebo acupuncture

Inert placebo treatment is crucial to RCTs. It makes the participants believe that they are receiving the correct treatment, thereby eliminating the placebo effect of the intervention. Being distinguishable from real treatment is the greatest feature of a "placebo." Ethical issues may arise with the use of placebo acupuncture in RCTs in some circumstances; they are not always harmless. Methodological difficulties are encountered in selecting appropriate controls in randomized controlled acupuncture trials. Distinguishing between the specific and non-specific effects of acupuncture through sham control seems to be impossible. Even so, sham acupuncture is still identified as the process of controlling acupuncture treatment to blind participants and control non-specific placebo effects.

Mimicking acupuncture without physiological activity

is difficult, because too many factors are unrecognizable. According to a report from International Acupuncture Research Forum [20], the three important elements of placebo acupuncture are described as follows: (1) the effect of needle stimulation, in which the placebos need to be applied at the correct sites; (2) the effect of acupuncture points in general, in which the placebos should be applied off-site and off-meridian; and (3) the effect of the specific acupuncture points for a patient or condition, in which the placebos need to be applied at irrelevant acupuncture points. Placebo or sham acupuncture must meet two principles, as follows: (1) no difference in needle appearance, patient perception, and operation compared with real acupuncture; and (2) no positive effects. Sham acupuncture can be considered as an ineffective placebo only if participants are kept unaware of differences in acupuncture processes. The difficulty in developing ideal sham acupuncture while avoiding all active components in acupuncture practice is enormous.

Development and deficiency of sham acupuncture

The first report on control acupuncture involved the evaluation of acupuncture effect on knee arthritis in the 1970s; the researchers adopted a placebo acupoint [28]. This kind of sham acupuncture control is not reasonable and is designed based on the principle that the acupuncture points are specific factors of acupuncture.

The famous placebo acupuncture design was developed by Streitberger [29]. With characteristics of blunt tip, retractable needle hand, highly simulated needling sensation, and no penetration, the Streitberger acupuncture was supposed to be an ideal control, especially when compared with previous non-acupoint designs. However, the credibility and normalization of the design were questioned in a study [30]. Some of the subjects (40%) can identify the difference between true acupuncture and sham acupuncture. Moreover, for patients with real acupuncture experience and De Qi sensation, this sham control and blind method is nearly inoperative. To distinguish whether the blunted tip sham acupuncture (BT) and round tip sham acupuncture (RT) are different, some researchers investigated the De Qi sensation meticulously; the need for sensation between two group varied [31]. These results indicated that sham acupuncture is difficult to achieve consistently and is highly dependent on the patient's perception.

Considering the shortcomings of non-penetrating acupuncture, a method of "minimal acupuncture" was developed. It involved trying to blind the patient by penetrating the skin as a real needle but only on the surface of acupuncture points. A famous German acupuncture clinical trial (GERAC) [32] adopted this sham acupuncture design in a study comprising 1162 patients with CLBP. No statistically significant difference was found for the real vs.

sham acupuncture ($P = 0.39$), real vs. conventional therapy ($P < 0.001$), and sham vs. conventional therapy ($P < 0.001$). No difference was found between true acupuncture groups, even if the treatments the participants were subjected to were significantly different from the conventional treatment. This ironic conclusion shows that "minimal acupuncture" might not be "inert." A subsequent study had confirmed the presence of a statistical difference in the De Qi perception between the patients in the superficial needle insertion group and those in the mock deep penetration group [33]. The depth of acupuncture is not the main specific factor for acupuncture.

Studies on mechanisms have indicated that the depth and rotation of acupuncture is associated with needling sensation and pressure pain threshold [34]. Regarding the "minimal acupuncture" stimulations, sensory pathways that densely innervate the skin's epidermis exist, and these pathways are likely activated by this "inert" needling. If the sensory pathways from cutaneous and deep tissues have redundant roles in mediating a treatment, then the current sham control will have issues. Some studies suggested that real acupuncture effects may share similar genetic, neurohumoral, and brain mechanisms with placebo effects [35–37]. When a needle is inserted into a designated point on the body and subjected to various stimuli, it activates a variety of nerves and neuroactive components [38]. Hall *et al.* reviewed placebo studies and RCTs and identified genomic effects on placebo response [39]. Attempting to peel off the placebo effect of acupuncture through a sham needling design is almost impossible.

Electroacupuncture (EA) has been widely compared with sham EA in recent RCTs due to its precise, repeatable, and standardized stimulation intensity and duration, as well as its simple and verifiable electrical parameters [40,41]. Correspondingly, the sham EA also faces great challenges. In a systematic review of sham EA [42], 17 kinds of sham EA methods have been identified from 94 RCTs, but only 24 RCTs reported the positive credibility of sham EA design. Many of the sham EAs adopted no electrical stimulation, no penetration, and therapeutic acupoints. Some sham designs integrated the design philosophy of Streitberger acupuncture.

Based on the common types of sham acupuncture design, we summarize four main domains, namely, location, stimulation, needling, and penetration in Table 1. The basic assumptions for choosing sham acupuncture are listed. It is convenient for us to select sham acupuncture reasonably and interpret the result of RCTs involving sham acupuncture correctly. Given the complexity of sham EA and its subtype, we did not integrate the sham EA.

Traps in sham acupuncture-based RCTs

According to Charlotte Paterson's and Paul Dieppe's study

Table 1 Common sham acupuncture design in RCTs

Domains	Experimental group	Control group	Underlying hypothesis of acupuncture effect
Location	Acupoint	Non-acupoint	Acupoint specific
	Meridian	Non-meridian	Meridian specific
Stimulation	De Qi sensation	Non-De Qi sensation	Perception specific
	Manual operation	Physical stimulation (electricity, magnet, laser, or visually virtual)	Stimulation specific (different parameter specific)
	Deep insert	Shallow insertion	Needling depth specific
Needling	Standard procedure	Non-standard procedure	Procedure specific
Penetration	Penetration skin	Non-penetration skin	Penetration specific

Note: Sham electroacupuncture is not included in the table.

[25], the factors of acupuncture intervention, include the diagnosis process, doctor–patient interaction, and patient perceptions. Their findings have important implications for the trial design of sham acupuncture controls. Most sham acupuncture treatments are based on the underlying hypothesis that needling is a typical element of acupuncture. The idea is that participants in the control group are subjected to everything except needling. However, complex and multi-element characteristics of acupuncture doom the design of sham acupuncture. Sham acupuncture passes these other characteristic elements to both groups. Therefore, the difference between the two groups may greatly underestimate the overall therapeutic effect of acupuncture. Such difference is used to distinguish between the physiological (specific) and psychological (non-specific) effects of acupuncture, but trials have shown that sham acupuncture has a psychological or non-specific effect, because it is better than conventional therapies used in routine care.

Sham acupuncture is inaccurately deemed to be as powerful as real acupuncture. In the sham acupuncture RCTs, the comparison of some parameters of acupuncture is considered rather than the full characteristic efficacy of acupuncture [43]. The sham acupuncture is only suitable for comparing two acupuncture interventions, for example, comparing the effects of different needle methods. Some scholars pointed out that standard treatments and individual responses should be considered before determining whether acupuncture is useful [44]. A study on 110 post-surgery pain patients implicated that patients' perception and expectations may contribute to self-reinforcing effects and enhance the efficacy of acupuncture analgesia [45]. Consequently, from a clinical perspective, the use of sham acupuncture to control nonspecific factors might be unnecessary.

A meta-analysis for musculoskeletal pain investigated the sham acupuncture types in 61 RCTs [46]. The subgroup analysis of low back pain indicated that compared with true acupuncture, the pain alleviation in different sham acupuncture types are as follows: -1.23 (-1.98 to -0.48) for non-penetration; -0.19 (-0.31 to

-0.08) for superficial penetration; and -0.50 (-0.85 to -0.14) for normal penetration. Despite the heterogeneity ($I^2 = 73.0\%$), findings still suggested that the most inert and most receivable sham acupuncture is a non-penetrating needle design. Therefore, when we use sham acupuncture as a control and conclude that no difference exists between real acupuncture and sham acupuncture, we should interpret the results with caution.

The interpretive RCT research model used to evaluate the true clinical effect of acupuncture may have flaws. Hence, we should be aware of the traps mentioned above. We have dissected the details of acupuncture and RCT model from different domains, as follows: intervention principle, intervention attributes, population, blind method, and control. In Table 2, acupuncture RCT models may show contrasting characteristics to those of acupuncture. By understanding the characteristics of acupuncture and RCT models, we can develop a suitable research paradigm for acupuncture.

Discussion

According to the abovementioned analysis, what seems to be controversial evidence of using acupuncture for CLBP treatment is a limitation of clinical evaluation methodology of acupuncture. Indiscreetly obeying the scientific framework of RCTs may lead to the loss of the characteristics and nature of acupuncture, to the deviation from the essence and significance of acupuncture research, and to the violation of the original intention of efficacy evaluation for optimizing clinical practice.

Placebo or sham acupuncture designs in RCTs are worth thinking. The overall curative effect of acupuncture is separated in this research model, which does not fit with the clinical practice of acupuncture. Inferring causation by simply evaluating the need rather than the implementation process is unwise and inappropriate. The obtained evidence on acupuncture may not be convincing and may need to be interpreted with caution. For the design of sham acupuncture, we cannot speculate on the

Table 2 Traps or flaws of acupuncture RCT models in different domains

Domains	Acupuncture characteristics	RCT model	Flaws or traps in acupuncture RCTs
Intervention principle	Individual and integration concept	Homogenization and standardization	Regression from optimal to average
Intervention attributes	Integration of specific and non-specific factors	The intervention itself is the only specific factor	The non-specific factors of acupuncture were dissected
Population	Wide crowd	Fixed population	Draws conclusions from limited population
Blind method	Operated by an acupuncturist without blinding	Strictly double blind or triple blind	The blind method has poor implement ability and low reliability
Control	Sham acupuncture or minimal acupuncture	Blank control or inert treatment	Control styles are “non-inert”
Acupuncture regimen	Flexible	Stationary	Ignores the correspondence between treatment and disease change
Treatment	Dynamic adjustment	Stable	Loss of acupuncture characteristics and flexibility
Effect composition	Multiple (acupoint, stimulus, operation, environment, and psychology and expectation)	Monotonous (needling or electric stimulus)	Acupuncture properties are split
Diagnosis and treatment sequence	Traditional Chinese medicine diagnosis: synchronous or dynamic changes	Biomedical diagnosis: diagnosis is fixed and made prior to treatment	Violates the concept of acupuncture treatment
Evaluation indicators	Subjective and objective indicators	Objective biological indicators	Ignores the patient’s subjective perception of treatment
Settings	Real-world	Ideal-world	Poor extrapolation

characteristics of the placebo control. We should conduct comprehensive internal and external reliability verifications of blind methods among the participants. Moreover, a pre-experiment is necessary before using the sham acupuncture design in RCTs. The acupuncture effect is related to multi-organ and multi-system integration. This effect is related to acupoints and to stimulation methods, stimulation amount, and human body status. Consequently, the efficient and realistic design of sham acupuncture requires innovation. A research model that accords with the characteristics of acupuncture needs to be further explored. The methodological challenges in RCTs of acupuncture is inevitable.

First, the appropriate choice of sham acupuncture is important. We can achieve a better design only with the clear comprehension of the mechanism underlying acupuncture. The development of sham acupuncture should be based on the holistic acupuncture model and inert verification. Penetrating sham acupuncture is possibly a defective control of acupuncture in RCTs and should be used prudently. Second, the applicability of the evaluation system needs to be considered. We recommend that the efficacy evaluation of acupuncture should be compared with standard treatment rather than a sham procedure. Of course, before the evaluation’s usefulness can be judged, the individual response to acupuncture needs to be confirmed by sham control [47]. From a broader perspective of therapeutic evaluation, promoting acupuncture’s effectiveness is important. Therefore, the design of

RCTs should be more pragmatic [48] and should involve a population with good knowledge of acupuncture, an acceptable standard treatment control, and a meaningful outcome selection. Sham-controlled RCTs have three arms, namely, acupuncture, sham acupuncture, and waitlist (will receive real acupuncture later). These three arms are dependable. Different from the current design of sham-controlled RCTs, this design can assess acupuncture effectiveness systematically and helps avoid the trap of “placebo effect.” This approach can integrate the explanatory (blinded) and pragmatic (unblinded) design. Third, the research strategy needs consideration. According to the characteristics of acupuncture, we advocate the basic strategy of “walking with two legs.” One method uses RCTs to conduct a confirmatory study on the net effect of acupuncture for CLBP. This method depends on mature acupuncture programs to obtain approved evidence for clinical guidelines and mechanism clarification. The other method includes meticulous studies on individualized characteristics of acupuncture and uses real-world clinical research paradigms such as registry study [24] to continuously optimize and deconstruct the clinical treatment process elements of acupuncture. Two research models can be combined to accumulate high-quality data and evidence for the development of acupuncture. Moreover, reporting the RCTs in accordance with CONSORT-STRICTA extension strictly [49] may contribute to the accurate interpretation and ready replication of acupuncture.

Table 3 Comparison of different clinical research methods in acupuncture

Elements	Efficacy	Effectiveness	Implementation
Hypothesis	Acupuncture is invalid	Acupuncture is not as effective	Reproducibility and extensibility of the acupuncture program are poor
Object	Strictly limited participants	Limited participants	Participants, implementation environment, and strategies
Intervention program	Single fixed acupuncture program	Single or multiple dynamic acupuncture programs	Acupuncture education and training (skills, knowledge, and others); single or multiple dynamic acupuncture programs
Control selection	Placebo or blank control	Positive control or standard treatment or waitlist	Health strategy or policy
Outcomes	Disease-related indicators	Health-related indicators	^a Implementation-related indicators and health-related indicators
Research purpose	Verification of the acupuncture effect	Comparison of the acupuncture effectiveness or program optimization	Reproducibility and extension of effective acupuncture programs
Validity	Internal $\geq$ external	External $\geq$ internal	Focus on applicability and fidelity
Model examples	Explanatory RCTs	^b CERs, pragmatic RCTs, and others	Implementation research design, hybrid design, and others.

^aAcceptability, applicability, feasibility, and others. ^bComparative effect research.

The most clinically important issue is not whether acupuncture works better than sham acupuncture; it is for whom and under what conditions acupuncture can work better. When the effect of the mechanism is unknown and clinical trials cannot be repeated, the assessment of the implementation of interventions in clinical practice should include an assessment of the implementation process and patient outcomes [50]. Moreover, researchers need to focus on the implementation of acupuncture rather than needling only. The acupuncturist is the integrator of acupuncture knowledge and experience. As the implementer of acupuncture operation and the dominant power of acupuncture treatment, the wisdom of the acupuncturist is an integral part of the acupuncture process. Therefore, while evaluating and affirming the curative effect of acupuncture, the evaluation system of acupuncture implementation may shift in one direction in the future. We have compared the different clinical research methods used in acupuncture in Table 3, and such comparison may help us choose a suitable research design easily.

Conclusions

The controversy on the effectiveness of acupuncture for CLBP treatment is only a microcosm of the limitation of evaluation methodology. We need to be familiar with and try to avoid the pitfalls of the RCT model. Moreover, we need to focus on reasonable sham acupuncture. The task of establishing a new evaluation system that is in line with the clinical characteristics of acupuncture is arduous but promising. We believe that the amount of high-quality evidence that matches the characteristics of acupuncture will increase in the future.

Acknowledgements

This work was supported by the National Key Research and Development Program—International Cooperation Research on Evaluation of the Effect of Acupuncture on Superiority Diseases (No. 2017YFC1703600). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Compliance with ethics guidelines

Xuqiang Wei and Baoyan Liu declare that they have no conflict of interest. This manuscript does not involve a research protocol requiring approval by a relevant institutional review board or ethics committee.

References

1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Hoy D, Karppinen J, Pransky G, Sieper J, Smeets RJ, Underwood M, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. *Lancet* 2018; 391(10137): 2356–2367
2. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392(10159): 1859–1922
3. Salzberg LD, Manusov EG. Management options for patients with chronic back pain without an etiology. *Health Serv Insights* 2013; 6: 33–38
4. Buchbinder R, van Tulder M, Öberg B, Costa LM, Woolf A, Schoene M, Croft P, Buchbinder R, Hartvigsen J, Cherkin D, Foster NE, Maher CG, Underwood M, van Tulder M, Anema JR, Chou R,

- Cohen SP, Menezes Costa L, Croft P, Ferreira M, Ferreira PH, Fritz JM, Genevay S, Gross DP, Hancock MJ, Hoy D, Karppinen J, Koes BW, Kongsted A, Louw Q, Öberg B, Peul WC, Pransky G, Schoene M, Sieper J, Smeets RJ, Turner JA, Woolf A. Low back pain: a call for action. *Lancet* 2018; 391(10137): 2384–2388
5. Foster NE, Anema JR, Cherklin D, Chou R, Cohen SP, Gross DP, Ferreira PH, Fritz JM, Koes BW, Peul W, Turner JA, Maher CG; Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet* 2018; 391(10137): 2368–2383
 6. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CC, Chenot JF, van Tulder M, Koes BW. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J* 2018; 27(11): 2791–2803
 7. Qaseem A, Wilt TJ, McLean RM, Forciea MA; Clinical Guidelines Committee of the American College of Physicians. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2017; 166(7): 514–530
 8. Weiß J, Quante S, Xue F, Muehe R, Reuss-Borst M. Effectiveness and acceptance of acupuncture in patients with chronic low back pain: results of a prospective, randomized, controlled trial. *J Altern Complement Med* 2013; 19(12): 935–941
 9. Plunkett A, Beltran T, Haley C, Kurihara C, McCoart A, Chen L, Wilkinson I, Cohen SP. Acupuncture for the treatment of chronic pain in the military population: factors associated with treatment outcomes. *Clin J Pain* 2017; 33(10): 939–943
 10. Wang H, Yang G, Wang S, Zheng X, Zhang W, Li Y. The most commonly treated acupuncture indications in the United States: a cross-sectional study. *Am J Chin Med* 2018; 46(7): 1387–1419
 11. Vickers AJ, Vertosick EA, Lewith G, MacPherson H, Foster NE, Sherman KJ, Irnich D, Witt CM, Linde K; Acupuncture Trialists' Collaboration. Acupuncture for chronic pain: update of an individual patient data meta-analysis. *J Pain* 2018; 19(5): 455–474
 12. Skelly AC, Chou R, Dettori JR, Turner JA, Friedly JL, Rundell SD, Fu R, Brodt ED, Wasson N, Winter C, Ferguson AJR. AHRQ Comparative Effectiveness Reviews. Noninvasive nonpharmacological treatment for chronic pain: a systematic review. Rockville (MD): Agency for Healthcare Research and Quality (US), 2018
 13. Cho YJ, Song YK, Cha YY, Shin BC, Shin IH, Park HJ, Lee HS, Kim KW, Cho JH, Chung WS, Lee JH, Song MY. Acupuncture for chronic low back pain: a multicenter, randomized, patient-assessor blind, sham-controlled clinical trial. *Spine* 2013; 38(7): 549–557
 14. Bishop FL, Yardley L, Prescott P, Cooper C, Little P, Lewith GT. Psychological covariates of longitudinal changes in back-related disability in patients undergoing acupuncture. *Clin J Pain* 2015; 31(3): 254–264
 15. National Guideline Centre (UK). Low Back Pain and Sciatica in Over 16s: Assessment and Management. London: National Institute for Health and Care Excellence (UK), 2016
 16. Xu M, Yan S, Yin X, Li X, Gao S, Han R, Wei L, Luo W, Lei G. Acupuncture for chronic low back pain in long-term follow-up: a meta-analysis of 13 randomized controlled trials. *Am J Chin Med* 2013; 41(1): 1–19
 17. Wise J. NICE recommends exercise and not acupuncture for low back pain. *BMJ* 2016; 352: i1765
 18. Van Zundert J, Hans G, van Kuijk S, Van Boxem K, Vissers K. Low back pain. *Lancet* 2018; 392(10164): 2548–2549
 19. Leibling E, Leonhardt U, Köster 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): 189–196
 20. White AR, Filshie J, Cummings TM, International Acupuncture Research Forum. Clinical trials of acupuncture: consensus recommendations for optimal treatment, sham controls and blinding. *Complement Ther Med* 2001; 9(4): 237–245
 21. Berman BMBirch S, Cassidy CM, Ezzo J, Han JS, Lao LX, Oleson T, Pomeranz B, Shang C, Stux G, Takeshige C. *Clinical Acupuncture: Scientific Basis*. Berlin: Springer Science & Business Media, 2001
 22. Birch S. Clinical research on acupuncture. Part 2. Controlled clinical trials, an overview of their methods. *J Altern Complement Med* 2004; 10(3): 481–498
 23. Paterson C, Britten N. Acupuncture as a complex intervention: a holistic model. *J Altern Complement Med* 2004; 10(5): 791–801
 24. Wei X, Liu B, He L, Yang X, Zhou J, Zhao H, Liu J. Acupuncture therapy for chronic low back pain: protocol of a prospective, multicenter, registry study. *BMC Musculoskelet Disord* 2019; 20(1): 488
 25. Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ* 2005; 330(7501): 1202–1205
 26. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-world evidence—what is it and what can it tell us? *N Engl J Med* 2016; 375(23): 2293–2297
 27. Ernst E. Acupuncture—a critical analysis. *J Intern Med* 2006; 259(2): 125–137
 28. Gaw AC, Chang LW, Shaw L-C. Efficacy of acupuncture on osteoarthritic pain. A controlled, double-blind study. *N Engl J Med* 1975; 293(8): 375–378
 29. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet* 1998; 352(9125): 364–365
 30. 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): 401–409
 31. Younbyoung C, Yunju K, Ilhwan C, Sabina L, Sangjae L, Hyejung L. A comparison study of acupuncture sensation scale between real acupuncture and sham needle. *Korean J Acupunct* 2006; 23(4): 85–99
 32. Haake M, Müller HH, Schade-Brittinger C, Basler HD, Schäfer H, Maier C, Endres HG, Trampisch HJ, Molsberger A. German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. *Arch Intern Med* 2007; 167(17): 1892–1898
 33. Benham A, Phillips G, Johnson MI. An experimental study on the self-report of acupuncture needle sensation during deep needling with bi-directional rotation. *Acupunct Med* 2010; 28(1): 16–20
 34. Choi Y, Cho S, Lee J, Moon W, Yoo D. P02.116. The relationship between deqi and the effect of acupuncture. *BMC Complement Altern Med* 2012; 12(S1): P172
 35. Hall KT, Loscalzo J, Kaptchuk T. Pharmacogenomics and the placebo response. *ACS Chem Neurosci* 2018; 9(4): 633–635

36. Lewith GT, White PJ, Kaptchuk TJ. Developing a research strategy for acupuncture. *Clin J Pain* 2006; 22(7): 632–638
37. 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): 2456–2461
38. Zhang ZJ, Wang XM, McAlonan GM. Neural acupuncture unit: a new concept for interpreting effects and mechanisms of acupuncture. *Evid Based Complement Alternat Med* 2012; 2012: 429412
39. Hall KT, Loscalzo J, Kaptchuk TJ. Genetics and the placebo effect: the placebo. *Trends Mol Med* 2015; 21(5): 285–294
40. Liu Z, Liu Y, Xu H, He L, Chen Y, Fu L, Li N, Lu Y, Su T, Sun J, Wang J, Yue Z, Zhang W, Zhao J, Zhou Z, Wu J, Zhou K, Ai Y, Zhou J, Pang R, Wang Y, Qin Z, Yan S, Li H, Luo L, Liu B. Effect of electroacupuncture on urinary leakage among women with stress urinary incontinence: a randomized clinical trial. *JAMA* 2017; 317(24): 2493–2501
41. Zhang X, Wang Y, Wang Z, Wang C, Ding W, Liu Z. A randomized clinical trial comparing the effectiveness of electroacupuncture versus medium-frequency electrotherapy for discogenic sciatica. *Evid Based Complement Alternat Med* 2017; 2017: 9502718
42. Chen ZX, Li Y, Zhang XG, Chen S, Yang WT, Zheng XW, Zheng GQ. Sham electroacupuncture methods in randomized controlled trials. *Sci Rep* 2017; 7(1): 40837
43. Kim TH, Kang JW, Lee MS. What is lost in the acupuncture trial when using a sham intervention? *Acupunct Med* 2017; 35(5): 384–386
44. Lundeberg T, Lund I, Sing A, Näslund J. Is placebo acupuncture what it is intended to be? *Evid Based Complement Alternat Med* 2011; 2011: 932407
45. Vase L, Baram S, Takakura N, Yajima H, Takayama M, Kaptchuk TJ, Schou S, Jensen TS, Zachariae R, Svensson P. Specifying the nonspecific components of acupuncture analgesia. *Pain* 2013; 154(9): 1659–1667
46. Yuan QL, Wang P, Liu L, Sun F, Cai YS, Wu WT, Ye ML, Ma JT, Xu BB, Zhang YG. Acupuncture for musculoskeletal pain: a meta-analysis and meta-regression of sham-controlled randomized clinical trials. *Sci Rep* 2016; 6(1): 30675
47. Lund I, Näslund J, Lundeberg T. Minimal acupuncture is not a valid placebo control in randomised controlled trials of acupuncture: a physiologist's perspective. *Chin Med* 2009; 4(2): 1
48. Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016; 375(5): 454–463
49. MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, White A, Moher D; STRICTA Revision Group. Revised Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLoS Med* 2010; 7(6): e1000261
50. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health* 2003; 93(8): 1261–1267