Journal homepage:
www.jcimjournal.com/jim
www.elsevier.com/locate/issn/20954964
Available also online at www.sciencedirect.com.

• Commentary

The methodology flaws in Hinman's acupuncture clinical trial, Part II: Zelen design and effectiveness dilutions

Arthur Yin Fan

McLean Center for Complementary and Alternative Medicine, PLC., Vienna, VA 22182, USA

Keywords: acupuncture; chronic knee pain; Hinman; Zelen design; clinical trial; statistics; flaws

Citation: Fan AY. The methodology flaws in Hinman's acupuncture clinical trial, Part II: Zelen design and effectiveness dilutions. *J Integr Med*. 2015; 13(3): 136–139.

In the October 2014 publication of *JAMA*, Dr. Hinman and colleagues published the study "Acupuncture for Chronic Knee Pain: A Randomized Clinical Trial," which concluded that "in patients older than 50 years with moderate or severe chronic knee pain, neither laser nor needle acupuncture conferred benefit over sham for pain or function. Our findings do not support acupuncture for these patients[1]".

As pointed out in my former article, Part I[2], there were serious flaws in the trial design and statistics, as well as in the interpretation of the results. This article attempts to address problems in the Zelen design used by Hinman *et al*[1].

There are some advantages to using a Zelen design for a randomized controlled trial (RCT). First, a Zelen design has a post-randomization consent design, which means that consent is only sought for one treatment each time, without the uncertainty of randomization. Researchers can be more comfortable knowing that they have the participants' consent each time they undergo a treatment. Patients can also be more comfortable with this design because they know which type of treatment they are receiving; unlike traditional RCTs, patients are not ignorant of whether they are receiving the placebo or experimental treatment. Effects such as resentful demoralization and what is known as the "Hawthorne effect" (altered behavior or performance resulting from awareness of being a part of an experimental study[3]) become less of an issue as patients are not weary of being part of a new alternative group, only the "standard" therapy will applies to them[4,5].

However, it does have some disadvantages, and therefore can cause biases[4–6], which will be discussed in detail below.

1 High drop-out rate

In this trial[1], many patients dropped out from their original groups, particularly from the acupuncture, laser acupuncture, and sham laser acupuncture groups, which would bias the study results if these patients were not removed from the final data analysis.

Let us look over the original design: there were 71, 70, 71 and 70 patients in the control, acupuncture, laser acupuncture, and sham laser acupuncture groups at the beginning of the trial, respectively, for a total of 282 patients who started the study. However, there were only 69, 54, 58 and 54 patients (in the aforementioned groups, respectively) who actually completed the treatments at the end of the study or at week 12. The drop-out rates were 2.82% (2/71) in the control group; 22.86% (16/70) in the acupuncture group; 18.31% (13/71) in the laser acupuncture group; and 22.86% (16/70) for the sham laser acupuncture group. According to the acceptable standards for an RCT, drop-out rates less than 10% are acceptable, drop-out rates between

http://dx.doi.org/10.1016/S2095-4964(15)60172-8
Received February 20, 2015; **accepted** March 5, 2015.
Correspondence: Arthur Yin Fan, CMD, PhD, LAc; Tel: +1-703-499-4428; E-mail: ArthurFan@ChineseMedicineDoctor.US

10% and 20% mean that the resulting data quality is poor, and drop-out rates of more than 20% mean that the data quality is considered very poor and should not be used in analysis[7]. In this trial analysis, the data quality in the acupuncture and sham laser acupuncture groups are very poor as the drop-out rates are over 20%; the authors should not have directly used them in any statistical analysis, unless they had re-adjusted and re-balanced the sample among the groups during the study. As outlined by the National Institutes of Health[8], if there is a differential drop-out rate of 15% or higher between study arms, such as between the control group and the treatment group in this clinical trial, then there is a very high potential for bias. This is a flaw that can decrease the quality of the study results. And, the reasons causing the patients to drop out should be analyzed in detail. The large number of drop-outs probably reflects the hesitation of the patients enrolled in those groups to get those interventions, which also may cause patients who still remained in those groups to have under-valued the effects. For example, in the acupuncture group, the patients in that trial setting may under-value the effectiveness, compared to those in an actual clinical setting who would actively be choosing to try acupuncture.

2 The effectiveness in intervention groups was diluted by various factors

2.1 The data analyzed from the intervention groups included data from non-intervention patients

In Table 2 (page 1 317) of this paper[1], the stated number of patients still participating in the control, acupuncture, laser acupuncture, and sham laser acupuncture groups at the end of the study were 69, 64, 65 and 58, respectively. However, on page 1 315, the Figure shows that only 54 individuals in the acupuncture group, 58 in the laser acupuncture group, and 54 in the sham laser acupuncture group received all of the treatments and therefore completed the study. False negative data were kept in the intervention groups that diluted the effectiveness or results of the various interventions; in other words, some individuals placed in the intervention groups did not receive all of the treatments, or perhaps received no treatment at all. There were data from 10 participants (18.52%) in the acupuncture group, 7 participants (12.07%) in the laser acupuncture group, and 4 participants (6.90%) in the sham laser acupuncture group, for a total of 21 individuals, which were included in the final analysis, but were from patients who did not actually receive some or all of the specified interventions. This incorrect inclusion causes a significant type-II error, the false negative, due to the dilution effect. It is unclear whether the authors intentionally kept the inappropriate data (*i.e.*, data from patients who withdrew before completing the full course of treatments) as part of the original intervention groups while neglecting to note this in the write-up. Just how important this point is must be emphasized, as including data on participants who did not receive or complete treatments, and counting them in the analysis as if they were part of the intervention groups, can significantly impact the results of the analysis. The fact that this was not reported or mentioned in the study significantly affects the validity of the Hinman's study.

2.2 Zelen design itself caused the effectiveness dilution

2.2.1 Estimated dilution rates

The authors declined to give the original data for re-conducting statistical analysis, so the data published in the article[1] were used for dilution rate analysis in this article.

Let us expect that the actual effective rate (at week 12, estimated)[9] was 30% in the control group, at least 60% in both the laser acupuncture and acupuncture groups, and 40% in the sham laser acupuncture group. The crossover rates were 10.77% (7/65) in the laser acupuncture group, 6.90% (4/58) in the sham laser acupuncture group, and 15.63% (10/64) in the acupuncture group. If this is the case, using the calculation method described in the literature[5], the dilution rates should then be 21.87% in the laser acupuncture group, 13.80% in the sham laser acupuncture group, and 31.27% in the acupuncture group (the dilution rate calculations were shown in Tables 1–3). The dilution rate was very significant in the acupuncture group, which causes the effectiveness to be undervalued in the acupuncture group, by almost 1/3.

2.2.2 The effective significance was masked by limited sample size due to the Zelen design of this study

In a Zelen design RCT, the sample size will be much bigger than in a traditional RCT[4,5] due to patient drop-out from the original group(s), and patients being switched to other groups. For example, patients may be switched from the intervention group to the control group if they did not receive any treatment. In this particular Zelen RCT, the total number of patients that completed the study was actually 235. Originally, there were 282 participants in the start of the study; if we consider the patients that dropped out but still remained in the original group or the intervention group, we would have 256, and not 282 patients. This sample size is much smaller than that of many traditional RCTs[9–11], which makes the results of this study questionable. As there were patients that dropped out from the intervention groups, in which there were 10 patients from the acupuncture group, 7 from the laser acupuncture group, and 4 from the sham laser acupuncture group, such patients and their information should have been merged with the control group, or the group sizes should have been changed in the final analysis. According to calculations[5], when a traditional RCT needs 250 patients, if the crossover rate is 15% (or the dilution rate is 30%, which is similar to the data found in the acupuncture group), in a Zelen design RCT, the

Table 1 The dilution rate calculations of laser acupuncture group *vs* control group

Group	Intervention	Expected effective rate	Rate difference	Crossover rate*	ITT analysis	Rate difference	Diluting effect (%)
Control	None	30%		10.77%	(1−10.77%)×30%+ 10.77%×60%= 33.23%		
			30%			23.44%	(30%−23.44%)/ 30%=21.87
Laser acupuncture	Laser acupuncture	60%		10.77%	(1−10.77%)×60%+ 10.77%×30%= 56.77%		

*In laser acupuncture group, there were 7 patients switched or moved to control, 10.77%. ITT: intention to treat.

Table 2 The dilution rate calculations of sham laser acupuncture group *vs* control group

Group	Intervention	Expected effective rate	Rate difference	Crossover rate*	ITT analysis	Rate difference	Diluting effect (%)
Control	None	30%		6.90%	(1−6.90%)×30%+ 6.90%×40%=30.69%		
			10%			8.62%	(10%−8.62%)/10% =13.80
Sham laser acupuncture	Sham laser acupuncture	40%		6.90%	(1−6.90%)×40%+ 6.90%×30%=39.31%		

*In sham laser acupuncture group, there were 4 patients switched or moved to control, 6.90%. ITT: intention to treat.

Table 3 The dilution rate calculations of acupuncture group *vs* control group

Group	Intervention	Expected effective rate	Rate difference	Crossover rate*	ITT analysis	Rate difference	Diluting effect (%)
Control	None	30%		15.63%	(1−15.63%)×30%+ 15.63%×60%=34.69%		
			30%			20.62%	(30%−20.62%)/30% =31.27
Acupuncture	Acupuncture	60%		15.63%	(1−15.63%)×60%+ 15.63%×30%=55.31%		

*In acupuncture group, there were 10 patients switched or moved to control, 15.63%. ITT: intention to treat.

sample size will need to be 510 (more than double). It is standard that in larger RCTs for acupuncture studies the patient sample size is 500 to 800, some with even larger sample sizes[9–11]; using the Zelen design, therefore, the size should have been much bigger in order for this study to be valid.

3 The sample size calculation in this study is questionable

In general, multiple group RCTs need more vigorous statistical work. In the Hinman *et al*'s study[1], the calculation estimating sample size should have considered the characteristics of comparing among multiple groups, as well as the characteristics of both laser acupuncture and acupuncture. Although the authors did mention some factors in sample size calculation[12], the original major testing factor was laser acupuncture, even though the sample size calculation of this clinical trial was based on an article unrelated to either laser acupuncture or traditional acupuncture studies. This study's estimated simple size is based on an incorrect sample size calculation source. The sample size calculation should be based on either previously published studies on laser acupuncture or acupuncture, or if available, the authors' own previous studies in laser acupuncture or acupuncture.

The very small sample size is one of the key reasons that there were no significant differences between the laser

acupuncture group and the sham acupuncture group, or the laser acupuncture group and the acupuncture group in the statistical calculation.

4 Conclusion

The effectiveness of the acupuncture group was diluted 31.27%, and its drop-out rate was 22.86%, much higher than that of the other groups in Hinman's clinical trial, which constitutes major flaws in how this study is analyzed and interpreted[8]. Based on the bias of Zelen design used in the study, and incorrect sample size calculation, the conclusions drawn from this study are of poor quality, inaccurate, and invalid.

5 Acknowledgements

Dr. Arthur Yin Fan would like to thank Ms. Sarah Faggert for editing support.

6 Competing interests

The author is an independent researcher, and declares that he has no competing interests.

REFERENCES

1 Hinman RS, McCrory P, Pirotta M, Relf I, Forbes A, Crossley KM, Williamson E, Kyriakides M, Novy K, Metcalf BR, Harris A, Reddy P, Conaghan PG, Bennell KL. Acupuncture for chronic knee pain: a randomized clinical trial. *JAMA*. 2014; 312(13): 1313–1322.
2 Fan AY. The methodology flaws in Hinman's acupuncture clinical trial, Part I: Design and results interpretation. *J Integr Med*. 2015; 13(2): 65–68.
3 Campbell JP, Maxey VA, Watson WA. Hawthorne effect: implications for prehospital research. *Ann Emerg Med*. 1995; 26(5): 590–594.
4 Adamson J, Cockayne S, Puffer S, Torgerson DJ. Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemp Clin Trials*. 2006; 27(4): 305–319.
5 Ding Y. Zelen design (Randomized consent design/Prerandomization design). In: *The new advance on randomized control trial*. Page 23–35. [2015-02-22]. http://wenku.baidu.com/view/85ebd102581b6bd97f19ea08.html. Chinese.
6 Homer CS. Using the Zelen design in randomized controlled trials: debates and controversies. *J Adv Nurs*. 2002; 38(2): 200–207.
7 *Common-used statistical methods in evidence-based medicine*. [2015-02-22]. http://wenku.baidu.com/view/9b933612866fb84ae45c8de9.html. Chinese.
8 National Institutes of Health. Quality assessment of controlled intervention studies. *Systematic evidence reviews & clinical practice guidelines*. [2015-03-04]. http://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/rct.
9 Vickers AJ, Linde K. Acupuncture for chronic pain. *JAMA*. 2014; 311(9): 955–956.
10 Berman BM, Lao L, Langenberg P, Lee WL, Gilpin AM, Hochberg MC. Effectiveness of acupuncture as adjunctive therapy in osteoarthritis of the knee: a randomized, controlled trial. *Ann Intern Med*. 2004; 141(12): 901–910.
11 White A, Cummings M. Acupuncture for knee osteoarthritis: study by Hinman *et al* represents missed opportunities. *Acupunct Med*. 2015; 33(1): 84–86.
12 Hinman RS, McCrory P, Pirotta M, Relf I, Crossley KM, Reddy P, Forbes A, Harris A, Metcalf BR, Kyriakides M, Novy K, Bennell KL. Efficacy of acupuncture for chronic knee pain: protocol for a randomized controlled trial using a Zelen design. *BMC Complement Altern Med*. 2012; 12: 161.