DEVELOPMENT AND VALIDATION OF AN INDEX OF THE ADEQUACY OF THE CONTROL TECHNIQUE FOR ACUPUNCTURE STUDIES

DESENVOLVIMENTO E VALIDAÇÃO DE UM ÍNDICE DA ADEQUABILIDADE DA TÉCNICA DE CONTROLE PARA ESTUDOS DE ACUPUNTURA

David Gonçalves Nordon[1], Reinaldo José Gianini[2], Gisele Regina de Azevedo[3]

ABSTRACT

Background: unspecific effects of a technique may occur for several different reasons and play an important role when defining its efficacy. There is no information on any study in the Western world that has created or validated a scale that evaluates the possibility of the control technique exerting physiological effects and, therefore, underestimating the effect of acupuncture. Methods: a literature review was performed, and the most important items concerning the control technique that might exert a physiological effect were listed and scored according to their importance, based on the Jadad scale. Three experienced acupuncturists evaluated the index, suggested modifications or enhancements, and tested it. Results: score ranges from 0 to 10, from items 1 to 4, adding up to a total; next, such total is multiplied according to the importance attributed to it, on item 5, ranging the corrected total from 0 to 20. Despite the index's objectiveness, the specialists made some mistakes when using it, making the Friedman test non-significant. Discussion: the index was developed so that, the higher the score, the higher the probability of the control technique exerting a physiological effect. Despite the difference of scores among evaluators, these mistakes are rarely attributable to inter-rater subjectivity or uncertainties to items, as all of them are quite objective. Only one item in 145 scored showed diverging opinions. Conclusion: we have created an index of control adequacy (placebo or sham) for acupuncture studies that may be used with great objectivity by specialists and non-specialists.

Key-words: acupuncture, placebo, research methodology.

RESUMO

Introdução: efeitos inespecíficos de uma técnica podem ocorrer por diversas razões diferentes e têm um papel importante para definir a sua eficácia. Não há informação sobre nenhum estudo no mundo ocidental que tenha criado ou validado uma escala que avalie a possibilidade de uma técnica de controle exercer efeitos fisiológicos e, assim, subestimar o efeito da acupuntura. Métodos: uma revisão de literatura foi realizada e os itens mais importantes com relação à técnica de controle que poderiam exercer um efeito fisiológico foram listados e atribuídos uma nota de acordo com a sua importância, baseado na escala de Jadad. Três acupunturistas com experiência clínica avaliaram o índice, sugeriram modificações ou melhorias e o testaram. Resultados: o índice varia de 0 a 10, dos itens 1 a 4, somando um total; em seguida, tal total é multiplicado de acordo com a importância atribuída a ele, no item 5, chegando a um total corrigido de 0 a 20. Apesar da objetividade do índice, os especialistas cometeram alguns erros ao aplicá-lo, tornando o teste de Friedman não significativo. Discussão: o índice foi desenvolvido para que, quanto maior o escore, maior a probabilidade de a técnica de controle exercer um efeito fisiológico. Apesar de diferença de escores entre avaliadores, estes erros são dificilmente atribuíveis à subjetividade interavaliadores ou incerteza de itens, já que todos são bastante objetivos. Apenas em um entre 145 mostrou divergência de opiniões. Conclusão: nós criamos um item de adequabilidade do controle (Placebo ou Sham) para estudos de acupuntura que pode ser usado com grande objetividade por especialistas e não especialistas.

Descritores: acupuntura, placebo, metodologia da pesquisa.

BACKGROUND

Proving the efficacy of an intervention in medicine requires an adequate control to exclude the unspecific effects from the total effects of the intervention. Studies that focus on this are known as Randomized Controlled Trials (RCT) controlled by placebo and are currently considered as Level I Studies in Evidence-Based Medicine.[1] The unspecific effects of a technique may occur for several different reasons: firstly, for the Hawthorne effect, according to which a patient gets better just for being treated (i.e., receiving attention). This is as hard to control as is the regression to the mean which naturally occurs during the course of a disease. Not only that, effects related to the patient's expectations and contact with the doctor may enhance or worsen an intervention's effects.[1-4]

For acupuncture, all of those unspecific effects are important; therefore, the methodology of the control technique must be very accurate and precisely chosen for each type of study and each type of effect. However, heterogeneous controls that may exert physiological effects are commonly found in studies. Simply inserting a needle may release endorphins throughout the body and thus lead to analgesia.[5]

Nevertheless, several studies (whose control may have often over or underestimated the efficacy of the acupuncture technique) are included in meta-analyses (MA) regardless of this heterogeneity. And despite the efforts from several researchers to elaborate scales and indexes for the evaluation of methodology for other health related areas,[6] and even to the elaboration of guides or checklists for performing acupuncture studies[5] there is no information on any study in the Western

Rev. Fac. Ciênc. Méd. Sorocaba, v. 15, n. 1, p. 186 - 191, 2013

1. Acupunturista e shiatsuterapeuta pelo Colégio Brasileiro de Acupuntura; médico pela Pontifícia Universidade Católica de São Paulo.
2. Professor do Depto. de Medicina - FCMS/PUC-SP
3. Professora do Depto. de Ciências Fisiológicas - FCMS/PUC-SP

Recebido em 25/1/2012. Aceito para publicação em 13/9/2012.

Contato: d-nordon@uol.com.br

world that has created or validated a scale that evaluates the possibility of the control technique exerting physiological effects and, therefore, having underestimated the effect of the intervention.

Thus, this study intends to present an index that evaluates the adequacy of the control used in acupuncture RCTs.

METHODOLOGY

This study was approved by the Committee of Ethics on Research from Faculdade de Ciências Médicas de da Saúde de Sorocaba, in 2011. It was performed a literature review including most consulted textbooks of acupuncture in Brazil,[11-13] articles of recommendations of specialists regarding the technique,[8,9,14,15] and articles on the physiological mechanisms of acupuncture[5,16-21] and controls and their types.[2,22-28]

The main items related to control and that may modify its effects by producing a physiological effect were listed; to each one of them a score was given, representing its influence in the treatment and being proportional to its theoretical potential of physiological effect. Scoring was based mainly on the Jadad scale,[29] according to which, if there is no information on the item in the text, it should receive the worst score.

Next the index was evaluated by three acupuncturists with clinical experience, who could make suggestions on the items and scores. The acupuncturists tested the index in five different articles, in order to be familiarized with it and solve possible questions, and then applied it to eight different articles (one of them with two evaluable control groups), in order to analyze the inter-rater concordance. For each article, only information concerning the methodology was extracted; there was no mention to authors, journal, country or year of publication or any outcome or result, as to avoid possible biases.

RESULTS

The complete index is in Figure 1; score ranges from 0 to 10, from items 1 to 4, adding up to a total; next, such total is multiplied according to the importance attributed to it, on item 5, ranging the corrected total from 0 to 20.

Figure 1. Index of control adequacy (placebo or sham)

Item	Value
1. Selected points:	
Are the same from intervention group	4
Are local points, near those from intervention, in case of local pain	3
Are different from intervention, however traditional points	2
Are different from intervention, non-traditional points, however near those from intervention	1
Are different from intervention, non-traditional points and not near those from intervention	0
2. Needle insertion:	
Yes, deep in general and/or shallow in local pain	2
Yes, shallow	1
No	0
3. De Qi *achievement:*	
Yes	3
No	0
4. Needle manipulation:	
Yes	1
No	0
Total:	0-10
5. Length of control therapy was similar or the same as the intervention +	
Score = 7	X 2
Score 3-6	X 1,5
Score 0-2	X 1
Corrected total:	0-20

INSTRUCTIONS FOR USING THE INDEX

Details on scoring each item

If there is no information on the item in the text, it should receive the maximum score. A traditional point is considered as a recognized point, either included in a meridian or as an extra-point. An *Ah-shi* point is considered as a tender (i.e., painful when palpated), non-traditional point.

Details on each item's score:

1. Selected points

Are the same from intervention group: if the points from control are the same as those selected for real intervention group, it must be scored 4. If there is no data on the points selected for control, it is also scored 4.

Are local points, near those from intervention, in case of local pain: if the points chosen for control, when treating local pain, are near those from the intervention group, no matter what kind of point (traditional or *Ah-shi*), it must be scored 3, for their potential of influencing the physiological mechanisms of local pain, given its location, the relatively wide area of action from the traditional point and the obvious effects of an *Ah-shi* point.

Are different from intervention, however traditional points: if the points chosen for control are different from the intervention group, however still traditional points, even if not related to the pattern in treatment, score 2, as it may be treating a Chinese pattern in an unexpected way. The only exception is when the researchers affirm in the text having purposely chosen such points, as they have no proved effect on the pattern and would not influence the treatment in any way. Then, it should be scored 0.

Are different from intervention, non-traditional points, however near those from intervention: if the control points are non-traditional, different from the intervention ones and however near these, it must be scored 1. A point must be considered near if it is located within an area of 13cm^2 around the original point.[30]

Are different from intervention, non-traditional points and not near those from intervention: the score is 0 if the points from control are non-traditional, different from those of intervention and not near them, and not near *Ah-shi* points, in case of local pain.

2. Needle insertion

*Yes, deep in general and/or shallow in local pain: w*hen it is informed that the control needle was inserted deeply for treatments in general, and/or shallow for local pain treatment, this item must be scored 2. If there is no information on the depth of insertion, it must be also scored 2.

Yes, shallow: when needle insertion was only deep enough to keep it attached to skin, score 1.

No: if the needle has not pierced the skin, score 0.

3. *De Qi* achievement

Yes: this item must be scored 3 either if *De Qi* was achieved during needle insertion, or if there is no information on that in the text, for it must be controlled for, as *De Qi* means physiological action.

No: if there was no *De Qi* achievement through needle insertion, score 0. Also, if there was no skin piercing, there could be no *De Qi*, and thus 0 must be scored.

4. Needle manipulation

Yes: regardless of the technique (method or frequency) of manipulation, if there was any, it must be scored 1. Also, if there was no information on needle manipulation for control, it must be scored 1.

No: if there was no needle manipulation, score 0.

5. Length of control therapy was similar or the same as the intervention

A control that lasts as long as the intervention may eventually become physiologically effective, due to continuous stimuli, according to the variables analyzed in this index. Therefore, the total must be multiplied according to its value, as shown below:

Score≥7: if the length of control treatment was similar to the intervention group and the total score is of seven or higher, the result must be multiplied by 2, due to the great possibility of a physiological effect in the control group. Seven was chosen as a cut-out score, for it would indicate since from a deep insertion, with *De Qi* achievement and manipulation, in non-traditional, different from intervention and nearby points, or even the same points as intervention, with deep insertion and manipulation, or the same points as intervention, with *De Qi* achievement, up to deep insertion with *De Qi* achievement and manipulation in the same points as intervention. Such a wide range of possibilities allows a great chance of similar physiological effect in the control group.

Score 3-6: if the length of control therapy was similar or the same as the intervention and the score ranges from three to six, the result must be multiplied by 1.5. Such a large specter of possibilities, since from *De Qi* achievement in any point up to *De Qi* achievement through deep insertion with needle manipulation, or *De Qi* achievement in local points near those from intervention in local pain, would guarantee a moderate possibility of physiological effects in the control group.

Score 0-2: in case the score ranges from zero to two, it must be multiplied by one, as to maintain the same score, due to the small possibility of the control having a physiological effect.

Corrected total

Is the total possible score for the index. Ranges from zero to twenty in three different groups: 0-2, 4.5-9 and 14-20.

Statistical analysis

Despite the index's objectiveness, the specialists made some mistakes when using it, making the Friedman test non-significant. Therefore, it was chosen to reproduce the answers from the specialists and analyze what went wrong (Table 1 and 2). Evaluator 1 was considered as the gold-standard, as he followed all the recommendations from the instructions described above, and will be used as comparison for the other specialists.

Table 1. Scores of each article according to evaluator and item

Item	Evaluator	Art1	Art2	Art3	Art4	Art5	Art6	Art7	Art 8a	Art 8b
1	1	4	4	4	0	0	4	4	4	0
	2	4	4	4	2	1 [a]	4	4	1 [a]	2 [a]
	3	0 [a]	4	1 [b]	2 [c]	4 [a]	4	4	4	2 [a]
2	1	1	0	1	0	1	0	0	0	2
	2	2 [a]	0	1	0	1	0	2 [a]	0	0 [a]
	3	0 [a]	0	1 [a]	0	0 [a]	0	0	0	0 [b]
3	1	3	0	3	0	0	0	0	0	3
	2	3	3 [d]	3	3 [d]	3 [a]	3 [d]	3 [d]	0	0 [a]
	3	0 [a]	0	0 [b]	0	0	0	0	0	0 [b]
4	1	0	0	0	0	0	0	0	0	1
	2	1 [a]	0	1 [a]	0	1 [a]	1 [a]	1 [a]	1 [a]	0 [a]
	3	0	0	0	0	1 [a]	1 [a]	0	0	0 [b]
5	1	16	6	7.5	0	1	6	6	6	9
	2	20	14	18	7.5	9	16	20	2	2
	3	0	6	2	2	7.5	7.5	6	6	2

Subtitle: a: Reading mistake. b: Non-observation of the rules for fulfilling the index. c: Diverging opinions. d: Internally illogical.

Table 2. Number of mistakes when applying the index according to each evaluator

Mistake (total number/% of analyzed items)/Evaluator	2	3
Reading mistakes	15/33.3%	9/20%
Non-observation of rules	0	5/11%
Diverging opinions	0	1/2.2%
Internally illogical	4/8.8%	0

Evaluator 2 presented 19 mistakes, 15 (79%) of them attributable to reading comprehension of the original article. Internally illogical mistakes happened when the evaluator scored maximum for *De Qi* achievement (3) when there was no needle penetration. It is an illogical rather than a reading mistake, for even if it were mentioned in the text whether there was *De QI* achievement or not, it would be impossible to achieve it without needle penetration.

Evaluator 3 made 16 mistakes, nine (56,3%) of which related to reading comprehension, five (31,3%) for not observing the rules of application (if an item is not informed in the text it must be scored the highest) and one of diverging opinions, concerning the proximity of points used in article 4.

The specialist believed that a distance of 0.5 to 1 cm represented a different and nearby point, rather than the same acupoint, which represented a different score.

DISCUSSION

This study has used current available knowledge on the mechanisms of acupuncture and placebo and sham controls to elaborate an index capable of representing the possibility of physiological effects in the control technique. It is a very objective index, based exclusively on information from the article and relatively easy to apply. It may be used as well when elaborating the RCT, to evaluate the control and avoid confusions in the study. Items regarding point localization were listed and scored gradually, according to its similarities to the intervention therapy. The more similar, the higher the chance of influencing the study.

The area of point proximity was not randomly chosen; a study by Aird, 2002, has observed that the methods commonly used for point localization, such as hands measures (either proportionality or *cuns*), create an area of approximately 13 cm^2 for the point, what could be reduced to only 3 cm^2 by the use of a plastic ruler or measuring elastic bands. Once most of the physicians prefer using their hands than these two techniques (for, as they say, they are not "well seen" by the patients), it is considered that in the daily practice a point has a circular area of an approximately 2 cm radius. Therefore, it seems acceptable that this proximity still exerts some physiological effect. Thus, any point within this area may still exert specific, rather than unspecific, effects.

Corrected total is a way to attribute importance to the length of treatment; the longer a control with high physiological effect lasts, the higher the possibility of it working as the intervention. On the other hand, the lower physiological effect it exerts, no matter how similar the length of treatment is, the less likely it is to have an effect.

The great amplitude of scores that this correcting weight may bring has made the Friedman test non-significant for the total score. However disheartening such a difference of scores among evaluators may seem, when each item is separately analyzed, it is evident that those mistakes are rarely attributable to inter-rater subjectivity or uncertainties to items, as all of them are quite objective. Only one item in 145 showed diverging opinions, however very plausible, which may be easily solved once one considers, as exposed above, that a point's area has a 2 - centimeter radius. According to this, the points used in article 4 for the control would be the same from the intervention, rather than nearby points.

Nevertheless, it is important to consider that such area may vary from point to point, according to its localization. Its original study was performed on the forearm, for points depending on measures.

Points localized in defined and identifiable anatomical areas generally present little or no difference in localization by different therapists. In such a case, a 0.5 to 1 cm distance may not be, in fact, respective to the chosen point.

This index was thus elaborated with the intention that, the higher the score, the higher the possibility of the control having a physiological effect similar to the intervention. The cut-out scores are based on TCM theories and scientific evidences from the mechanisms of acupuncture and control procedures. As this index disseminates and is correlated to RCT, it will be possible to establish cut-out scores that may foresee the efficacy of acupuncture in the study, according to the physiological effect exerted by the control.

Limitations

This index was created for placebo and sham controls for acupuncture. It may be adapted in order to include electroacupuncture, in which case needle manipulation must be considered as presence of electrical stimulus. It was not conceived, though, to be used for TENS (Transcutaneous Electrical Nerve Stimulation), PENS (Percutaneous Electrical Nerve Stimulation) or other control techniques that are not, in fact, the most adequate for acupuncture RCT. Such limitation is, however, temporary; as this index disseminates, it will soon be able to be adapted to other techniques, following the same philosophy used in its conception.

CONCLUSION

We have created an index of control adequacy (placebo or sham) for acupuncture studies that may be used with great objectivity by specialists and non-specialists.

The authors recommend that it is not restricted only to systematic reviews and MA, but also that researchers consider it as a checklist to be followed when preparing a RCT, and that physicians use it before making a decision on a conduct, when considering the scientific information available.

Acknowledgements

We are thankful for the acupuncturists who helped us throughout this study: Drs. Carla A D Antonia and Viviane S P Giovaninni.

REFERENCES

1. Fletcher RH, Fletcher SW, Wagner EH. Epidemiologia clínica: elementos essenciais. 3ª ed. Porto Alegre: Artes Médicas; 1996.
2. Benedetti F. Mechanisms of placebo and placebo-related effects across diseases and treatments. Annu Rev Pharmacol Toxicol. 2008; 48:33-60.
3. Pariente J, White P, Frackowiak RSJ, Lewith G. Expectancy and belief modulate the neuronal substrates of pain treated by acupuncture. Neuroimage. 2005; 25:1161-7.
4. Wallach H. The efficacy paradox in randomized controlled trials of CAM and elsewhere: beware of the placebo trap. J Altern Complement Med. 2001; 7(3):213-8.
5. MacPherson H, Green G, Nevado A, Lythgoe MF, Lewith G, Devlin R, et al. Brain imaging of acupuncture: comparing superficial with deep needling. Neurosci Lett. 2008; 434:144-9.
6. Moher D, Jadad AR, Nichol G, Pernman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995; 16:62-73.
7. Birch S. Clinical research on acupuncture: part 2. Controlled clinical trials, an overview of their methods. J Altern Complement Med. 2004; 10(3):481-98.
8. Birch S. Issues to consider in determining an adequate treatment in a clinical trial of acupuncture. Complement Ther Med. 1997; 5:8-12.
9. Claraco AE, Fargas-Bablak A, Hanna SE. The Reporting of clinical acupunture research: what do clinicians need to know? J Altern Complementa Med. 2003; 9(1):143-9.
10. Hammerschlag R. Methodological and ethical issues in clinical trials of Acupuncture. The Journal of Alternative and Complementary Medicine, 1998; 4(2):159-71.
11. MacPherson H, White A, Cummings M, Jobst J, Rose K, Niemtzow R. Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. Complement Ther Med. 2001; 9:246-9.
12. Bensky D, O'Connor J; Shanghai College of Traditional Medicine. Acupuntura: um texto compreensível. São Paulo: Roca; 1996.
13. Maciocia G. Os Fundamentos da medicina chinesa: um texto abrangente para acupunturistas e fitoterapeutas. 2ª ed. São Paulo: Roca; 2007.
14. Birch S. Systematic reviews of acupuncture – are there problems with these? Clin Acupunct Oriental Med. 2001; 2:17-22.
15. MacPherson H, White A, Cummings M, Jobst J, Rose K, Niemtzow R. Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. Complement Ther Med. 2001; 9:246-9.
16. Asghar AUR, Green G, Lythgoe MF, Lewith G, MacPherson H. Acupuncture needling sensation: The neural correlates of deqi using fMRI. Brain Res. 2010; 1315:111-8.
17. Bäcker M, Hammes MG, Valet M, Deppe M, Conrad B, Tölle TR, Dobos G. Different modes of manual acupuncture stimulation differentially modulate cerebral blood flow velocity, arterial blood pressure and heart rate in human subjects. Neurosci Lett. 2002; 333:203-6.
18. Hamza MA, Ghoname EA, White PF, Craig WF, Ahmed HE, Gajraj NM, et al. Effect of the duration of electrical stimulation on the analgesic response in patients with low back pain. Anesthesiology. 1999;91:1622-7.
19. Kong J, MA L, Gollub RL, Wei J, Yang X, Li D, et al. A pilot study of functional magnetic resonance imaging of the brain during manual and electroacupuncture stimulation of acupuncture point LI-4 Hegu) in normal subjects reveals differential brain activation between methods. J Altern Complement Med. 2002; 8(4):411-9.
20. Napadow V, Dhond R, Park K, Kim J, Makris N, Kwong KK, et al. Time-variant fMRI activity in the brainstem and higher structures in response to acupuncture. Neuroimage. 2009; 47:289-301.
21. Zhang WT, Zhen J, Luo F, Zhang L, Zeng YW, Han JS. Evidence from brain imaging with fMRI supporting functional specificity of acupoints in humans. Neurosci Lett. 2004; 354:50-3.
22. Birch S. A review and analysis of placebo treatments, placebo effects, and placebo controls in trials of medical procedures when sham is not inert. J Altern Complement Med. 2006; 12(3):303-10.
23. Chae Y, Lee H, Kim H, Sohn H, Park JH, Park HJ. The neural substrates of verum acupuncture compared to non-penetrating placebo needle: An fMRI study. Neurosci Lett. 2009; 450:80-4.
24. Harris RE, Zubieta JK, Scott DJ, Napadow V, Gracely RH, Clauw DJ. Traditional Chinese acupuncture and placebo (sham) acupuncture are differentiated by their effects on μ-opioid receptors (MORs). Neuroimage. 2009; 47:1077-85.
25. Hróbjartsson A, Gotzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. J Intern Med. 2004; 256:91-100.

26. Lai X, Zhang G, Huang Y, Tang C, Yang J, Wang S, et al. A cerebral functional imaging study by positron emission tomography in healthy volunteers receiving true or sham acupuncture needling. Neurosci Lett. 2009; 452:194-9.
27. Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. Ann Intern Med, 2002; 136:471-6.
28. Wood R, Lewith G. The credibility of placebo controls in acupuncture studies. Complem Ther Med. 1998;6:79-82.
29. Jadad AR, Moore A, Carrol D, Jenkinson C, Reynolds DJM, Cavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clinical Trials, 1996;17:1-12.
30. Aird M, Cobbin DM, Rogers C. A study of the relative precision of acupoint location methods. J Altern Complementary Med. 2002; 8(5):635-42.