Article:

Webster, R., Bishop, F., Collins, G. et al. (2021) Measuring the success of blinding in placebo-controlled trials: should we be so quick to dismiss it? Journal of Clinical Epidemiology, 135. pp. 176-181. ISSN: 0895-4356

https://doi.org/10.1016/j.jclinepi.2021.02.022

Measuring the success of blinding in placebo-controlled trials: should we be so quick to dismiss it?

Running head: Measuring blinding success in placebo-controlled trials

Author list and affiliations

Rebecca K Webster*, PhD, University of Oxford, Oxford, United Kingdom and University of Sheffield, United Kingdom; Department of Psychology, Cathedral Court, 1 Vicar Lane, Sheffield, S1 2LT; r.k.webster@sheffield.ac.uk

Felicity Bishop, PhD, University of Southampton, Southampton, United Kingdom

Gary S Collins, PhD, University of Oxford, and NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom

Andrea WM Evers, PhD, Leiden University, Leiden, Netherlands

Tammy Hoffmann, PhD, Institute of Evidence-Based Healthcare, Bond University, Queensland, Australia

J. André Knottnerus, MD PhD, Maastricht University, Maastricht, Netherlands

Sarah E Lamb, DPhil, University of Oxford, Oxford, United Kingdom and University of Exeter, Exeter, United Kingdom

Helen Macdonald, MD, The BMJ, London, United Kingdom

Claire Madigan, PhD, University of Oxford, Oxford, United Kingdom

Vitaly Napadow, PhD, Harvard Medical School, Boston, United States

Amy Price, PhD, Stanford University, Stanford, United States; University of Oxford, Oxford, United Kingdom and The BMJ, London, United Kingdom,

Jonathan L Rees, MD, University of Oxford, and NIHR Oxford Biomedical Research Centre Oxford, United Kingdom

Jeremy Howick, PhD, University of Oxford, Oxford, United Kingdom

*Corresponding author

Word count: 2072

Key words: Blinding, masking, trials, measuring, reporting guidelines

1 Background

From being almost universally regarded as a methodological virtue of clinical trials and being included in the original 2001 Consolidated Standards of Reporting Trials (CONSORT) statement (1), measuring the success of blinding has fallen out of fashion. Subsequent versions of CONSORT removed this recommendation based on the correct view that it can lead to misleading inferences about causes of the failure to blind. (2, 3) In addition, Anand, et al. (4) recently questioned the need to blind patients and clinicians or measure and report whether blinding was done successfully. While critics are correct to point out problems with the view that blinding is a universal methodological virtue, and to point out that measuring the success of blinding is not straightforward, they are too quick to dismiss the value of testing and reporting on the success of blinding. This is reflected in our findings extending the Template for Intervention Description and Replication (TIDieR) statement for placebo/sham control components, in which almost all Delphi respondents recommended that trials should measure and report whether blinding was successful. (5)

We are not aware of any publications that set out the case for and against measuring blinding success, or that provide mitigating positions. Our experience suggests that confusion about blinding inhibits reasonable debates in this area. Here, we attempt to clarify some of the confusions surrounding blinding and measuring its success, before providing the case for and against, reporting measures of the success of blinding, and suggesting a 'middle road' which takes both sides of the debate into account.

2 Measuring blinding success: the case for

Blinding involves concealing knowledge of treatment assignment to one or more groups involved in clinical trials (participants, intervention providers, data collectors, outcome assessors, statisticians, and manuscript authors). (6) Trials can be described in a number of ways including open (unblinded), single-blind, double-blind or triple-blind. The terminology can be confusing however, as a random sample of 200 trials has shown that the term double blind can be used to describe blinding up to 18 different combinations of trial personnel. (7) As noted in CONSORT, it is important to specify who was blinded in a trial, (2) as blinding different people may affect outcomes, especially those which are subjective. For example, if participants and data collectors were not blinded this may have more of an impact than an unblinded statistician who may have less influence on the outcomes.

Measuring whether blinding was successful involves asking patients and clinicians about their treatment assignment beliefs before the trial is officially unblinded. Successful blinding occurs when there is a balance of expectations and beliefs related to the assigned intervention, demonstrating that those who are blinded are not aware of the (active or inactive) intervention that has been assigned. However, blinding can fail when participants, caregivers, or other groups involved in a trial deduce the intervention allocation at the beginning of the trial (e.g. due to inadequate matching between the placebo and active intervention), or during the trial (e.g. due to adverse events). (8-10) Since the function of blinding is to reduce the impact of expectations, unsuccessful blinding is problematic, as beliefs and expectations of those who correctly guess the intervention allocation *could* then influence the outcome of the trial. (11-14) As such a trial that was designed blinded but in which attempts to blind were unsuccessful may approach the quality of a trial where (complete, double) blinding is ethically and feasibly possible, but is not blinded (see Fig 1).

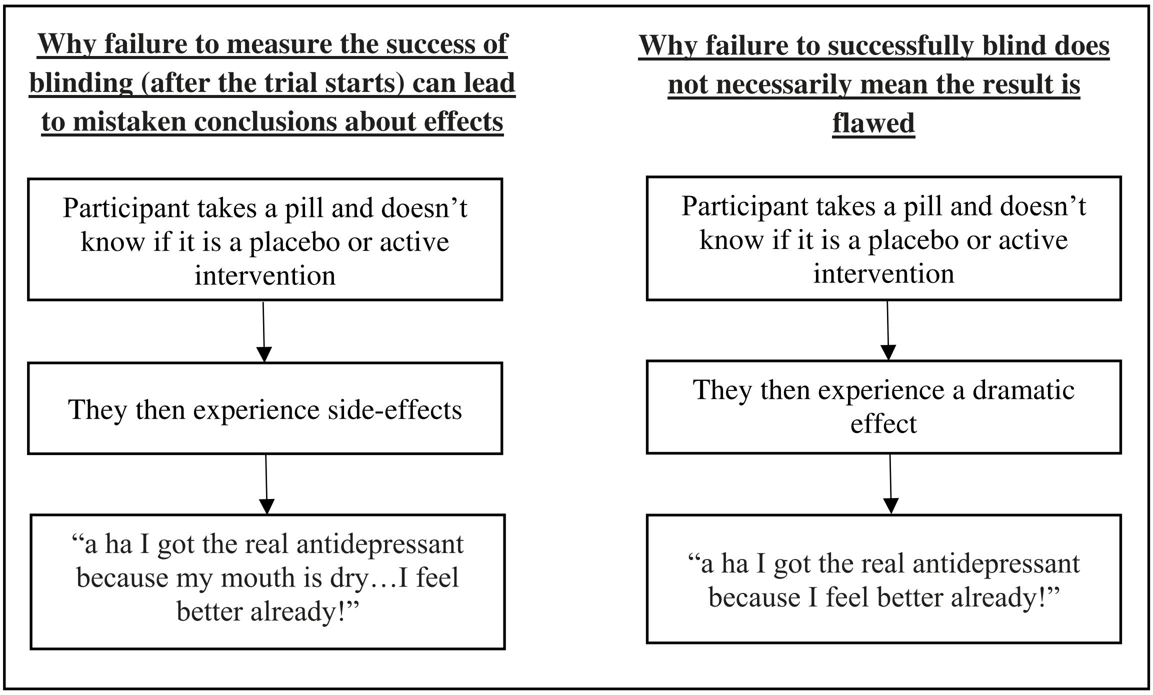

Fig 1. Why measuring blinding success is important and why it is not

A number of meta-epidemiological studies have investigated differences between trials (reported as) blinded and those that are not (reported as) blinded. (15-24) Some (but not all) of those found that lack of reporting of blinding led to larger effect sizes. Recently, Moustgaard, et al. (15) found inconsistent effects of blinding on treatment effect sizes. However, there are methodological concerns regarding the study's sample selection and classifications of reporting of blinding. (25) Like randomisation and allocation concealment, blinding can reasonably be expected to have a small average effect, possibly with an unpredictable direction. (26, 27) In an era when marginal gains from many of our medical interventions suffice to change policy and practice, (28) ruling out small biases or errors is becoming more important. In addition, small average effects are compatible with larger effects in some instances, for example trials of treatments for disorders that are placebo responsive, such as pain. Additional meta-epidemiological studies with large sample sizes, together with well-defined outcomes, disease areas, and classifications of reporting of blinding are required to address this important issue. Such studies cannot be conducted unless trials report whether blinding was successful (where this is feasible).

Aside from the importance of blinding itself, the importance of measuring (see Box 1) and reporting blinding success is apparent in various trials. For example, Karlowski, et al. (29) compared Vitamin C with placebo for treating the common cold, and found Vitamin C to be apparently effective. However, because of the sour taste of Vitamin C and sweet taste of the lactose placebo pills, the trial was not successfully blinded. When the authors carried out a subgroup analysis in which they divided participants into those who remained blinded and to those who were not, they found that there was no benefit of Vitamin C in the blinded group. Although ideally the authors should have ensured both placebo and active intervention were adequately matched, this example still shows the importance of measuring and reporting blinding success. Otherwise, it would have been mistakenly concluded that Vitamin C was superior.

More recently, a unsuccessfully blinded trial of zinc for treating common cold symptoms found that zinc significantly reduced the duration of cold symptoms compared to placebo. (30) Whereas, another trial with successful blinding, found that zinc did not reduce symptom duration. (31) This difference may be due to significantly more side-effects being reported to Zinc than placebo in the first trial, (30) which led to unblinding and subsequent bias. As such the success of blinding reported in these studies could be useful for those appraising them and looking for reasons for their discrepant results.

A common approach to measuring the success of blinding uses chi-square tests of independence, where successful blinding is indicated by a null finding (patient guesses are not related to their intervention allocation). (32) However, this lacks sensitivity and does not provide any directional information about the pattern of participant guesses. (33) James' (34) and Bang's (33) blinding index (BI) have addressed some of these concerns by asking participants to guess their intervention assignment using three responses (active, placebo or do not know). James' provides a single value that combines data from all arms ranging from 0 to 1, 0 being total lack of blinding, 1 being complete blinding and 0.5 being completely random blinding. Bang's BI aims to provide a more sensitive measure of blinding within each experimental arm compared to James' by calculating a score from -1 to 1, 1 being complete lack of blinding, 0 being consistent with perfect blinding and -1 indicating opposite guessing which may be related to unblinding. (33) As such, it can be used to detect where blinding may have failed, while still assessing overall success. An even newer method is the use of video surveillance. This involves video-recording procedures in the trial and asking a professional familiar with the procedure to guess the intervention allocation. (35) However, in practice, blinding success is rarely measured, with only 2-24% of trials reporting the success of blinding. (36, 37). In addition, these methods fall short as they do not consider why unblinding may have occurred.

Box 1. How to measure blinding success?

3 Measuring blinding success: the case against

The case against measuring the success of blinding can be traced to Dave Sackett, who cited a 2x2 factorial trial of aspirin and sulfinpyrazone for stroke prevention. In the trial, blinded clinicians largely distinguished aspirin from sulfinpyrazone. (38) But, because of prior 'hunches' that sulfinpyrazone would be more effective, they mistakenly believed that patients with better outcomes had received sulfinpyrazone, when in fact the trial showed aspirin was more effective. In this example, the results of tests for blinding can be ambiguous. Hence, Sackett and others following him argued that tests for the success of blinding should not be conducted.

Sackett is correct that in this example (and perhaps others like it), that the test for the success of blinding was confounded by mistaken beliefs about which intervention was effective (or a misattributed response to treatment). However, if these (mistaken) hunches about efficacy were *different* (unbalanced) in the intervention and control groups, then they could have confounded the study no matter how mistaken they were. Or, their beliefs were the same (balanced) across the groups, in which case there was no confounding (even if the beliefs were mistaken). Either way, the test for the success of blinding will reveal useful information, namely about whether expectations might have confounded the results.

There are some cases in which failure to successfully blind does not imply that the study was methodologically lacking. For example, a dramatically effective treatment can cause unblinding, however it should not lead us to conclude that a trial of the treatment was methodologically lacking. On the contrary, as Senn (39) argued: 'The whole point of a successful double-blind trial is that there should be unblinding through efficacy." The problem remains however, that if a trial reports that the cause of unblinding was dramatic effectiveness, a report of 'failed' blinding could mislead some into thinking the trial was less trustworthy.

Secondly, measuring the success of blinding at the wrong time (for example before follow-up or trial completion) may raise suspicion among participants and cause the problem it is intended to prevent. (40) (41)

Thirdly, some trials cannot feasibly or ethically be blinded, for example, non-drug interventions such as exercise, behavioural therapy and nutritional advice. (Aside: trials of these interventions can be rigorous by using other methodological tools to reduce bias (42), such as pre-registering trials, following a pre-specified analysis plan, adequate sample size and using randomisation, to reach the best achievable research practice.) Also, in some cases unblinding is an ethical requirement, for example due to hypothesized toxicity, and blinding itself could increase research waste, with some evidence indicating that patients are less likely to enrol in blinded trials. (4)

4 Discussion

Demanding that all trials attempt to use and measure the success of blinding is too strong because blinding is sometimes impossible, unethical, or misleading. Future research is required to determine how to best interpret findings from assessing the success of blinding. On the other hand, blinding has the potential to rule out bias, and failure to recommend that the success of blinding be reported when it is measured, seems like wilful withholding of information that potentially useful.

In addition, the change in the CONSORT recommendation from asking researchers to report on success of blinding (if measured) to not asking, seems to have been based on arguments that may deserve revisiting. Of course, the fact that CONSORT does not explicitly recommend reporting on the success of blinding does not prevent reviewers from reporting it. However, the fact that CONSORT sites a paper by Sackett as the reason for removing it, in which he claims that testing the success of blinding is a 'mug's game' could be interpreted as a reason to avoid reporting on the success of blinding.

Also, while measuring the success of blinding at many (or the wrong) points may cause some problem, this does not imply that measuring success of blinding at a single (roughly) correct point is not useful. Moreover, empirical research suggests that getting the 'correct' point may not be required. Rees, et al. (43) have shown that the difference between a six-point assessment of blinding success during a trial and a two-point model is not significant.

Overall, the fact that difficulties, ethical problems, or ambiguity in measuring its success does not imply that it should be given up altogether.

5 Conclusion and recommendation? A middle ground

While we acknowledge there are a dearth of studies that have investigated this issue, more definitive evidence can only come from studies that measure the success of blinding. We recognise that some trials cannot feasibly or ethically be blinded, but it is important that trials that *could have* introduced blinding and measured its success, are distinguished from trials that could not have. Our suggestion for a way forward considers the current state of evidence for and against measuring the success of blinding. We hope this stimulates further discussion, and that future iterations of CONSORT reflect on our arguments and revisits this issue.

We suggest that:

1. **Authors should make every attempt to match the placebo and active intervention to avoid unblinding at the start of the trial and subsequent research waste.**
2. **When authors have measured the success of blinding they should report the results**.
3. **Critical appraisers should consider reasons why unblinding may have arisen before condemning a trial as having a high risk of bias**, or if blinding success has not been reported, they should assess whether it is possible that blinding has been compromised.
4. **Future development of measures to assess the success of blinding should ask those intended to be blinded what their intervention allocation beliefs were and why**. This can help disentangle the reasons (dramatic effects or side-effects), although the reason may not always be known for sure.

Competing interests

Declarations of interest: none

Contributor statement

Rebecca K Webster: Conceptualization, Visualisation, Project administration, Writing – Original draft preparation; **Jeremy Howick**: Conceptualization, Supervision, Funding acquisition, Writing – Review & Editing; **Felicity Bishop, Gary S Collins, Andrea WM Evers, Tammy Hoffmann, André Knottnerus, Sarah E Lamb, Helen Macdonald, Claire**

Madigan, Vitaly Napadow, Amy Price, Jonathan L Rees: Conceptualization, Writing – Review & Editing.

Funding

This work was partly supported by the University of Oxford Humanities Division REF Support Fund provided funding for part of this project (awarded to JH and RW), a VICI grant from the Netherlands Organization for Scientific Research (NWO) (Number: 45316004), and a European Research Council Consolidator Grant (ERC2013-CoG-617700) (awarded to AWME). VN was supported by the National Institutes of Health, National Center for Complementary and Integrative Health (R01- AT007550, R61/R33-AT009306, P01-AT009965), and the National Institute of 16 Arthritis and Musculoskeletal and Skin Diseases (R01- AR064367). GSC was supported by the NIHR Biomedical Research Centre, Oxford and Cancer Research UK (grant C49297/A27294). TH is supported by a National Health and Medical Research Council of Australia Senior Research Fellowship. None of the funders played any role in the study

References

1. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet. 2001;357(9263):1191-4.
2. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
3. Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. Lancet. 2010;375(9721):1144-6.
4. Anand, Rohan, Norrie, John, Bradley, Judy M, McAuley, Danny F, Clarke, Mike. Fool's gold? Why blinded trials are not always best. BMJ. 2020;368:l6228.
5. Howick J, Webster R, Rees J, MacDonald H, Price A, Bishop F, et al. TIDieR-Placebo: checklist guide to reporting placebo and sham controls. PLoS Medicine. In press.
6. Questioning Double Blinding as a Universal Methodological Virtue of Clinical Trials: Resolving the Philip's Paradox. The Philosophy of Evidence-Based Medicine:63-79.
7. Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials. 2006;3(4):360-5.
8. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. Cochrane Database Syst Rev. 2004(1):Cd003012.
9. Bello S, Wei M, Hilden J, Hróbjartsson A. The matching quality of experimental and control interventions in blinded pharmacological randomised clinical trials: a methodological systematic review. BMC Medical Research Methodology. 2016;16(1):18.
10. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? Int J Epidemiol. 2007;36(3):664-5.
11. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet. 2002;359(9307):696-700.
12. Spanos NP, Burgess CA, Cross PA, MacLeod G. Hypnosis, reporting bias, and suggested negative hallucinations. J Abnorm Psychol. 1992;101(1):192-9.
13. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. Canadian Medical Association Journal. 2013;185(4):E201-E11.
14. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Canadian journal of surgery. Journal canadien de chirurgie. 2010;53(5):345-8.
15. Moustgaard H, Clayton GL, Jones HE, Boutron I, Jorgensen L, Laursen DRT, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ. 2020;368:l6802.
16. Page MJ, Higgins JP, Clayton G, Sterne JA, Hrobjartsson A, Savovic J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. PLoS One. 2016;11(7):e0159267.
17. Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. J Clin Epidemiol. 2016;77:24-37.
18. Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, da Costa BR, Flores-Mir C. Influence of blinding on treatment effect size estimate in randomized controlled trials of oral health interventions. BMC Med Res Methodol. 2018;18(1):42.
19. Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in Physical Therapy Trials and Its Association with Treatment Effects: A Meta-epidemiological Study. Am J Phys Med Rehabil. 2017;96(1):34-44.

20. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ. 2012;344:e1119.
21. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. CMAJ. 2013;185(4):E201-11.
22. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Rasmussen JV, Hilden J, et al. Observer bias in randomized clinical trials with time-to-event outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. Int J Epidemiol. 2014;43(3):937-48.
23. Hrobjartsson A, Emanuelsson F, Skou Thomsen AS, Hilden J, Brorson S. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. Int J Epidemiol. 2014;43(4):1272-83.
24. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials. Annals of internal medicine. 2012.
25. Howick J. Re: Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ. 2020;368:l6802.
26. Howick J, Mebius A. In search of justification for the unpredictability paradox. Trials. 2014;15:480.
27. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ. 1998;317(7167):1185-90.
28. Taylor F, Huffman MD, Macedo AF, Moore TH, Burke M, Davey Smith G, et al. Statins for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev. 2013;1:CD004816.
29. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic Acid for the Common Cold: A Prophylactic and Therapeutic Trial. JAMA. 1975;231(10):1038-42.
30. Prasad AS, Fitzgerald JT, Bao B, Beck FWJ, Chandrasekar PH. Duration of Symptoms and Plasma Cytokine Levels in Patients with the Common Cold Treated with Zinc Acetate: A Randomized, Double-Blind, Placebo-Controlled Trial. Annals of Internal Medicine. 2000;133(4):245-52.
31. Smith DS, Helzner EC, Nuttall CE, Jr., Collins M, Rofman BA, Ginsberg D, et al. Failure of zinc gluconate in treatment of acute upper respiratory tract infections. Antimicrob Agents Chemother. 1989;33(5):646-8.
32. Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. J Clin Epidemiol. 2005;58(12):1220-6.
33. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Controlled Clinical Trials. 2004;25(2):143-56.
34. James KE, Bloch DA, Lee KK, Kraemer HC, Fuller RK. An index for assessing blindness in a multi-centre clinical trial: disulfiram for alcohol cessation--a VA cooperative study. Stat Med. 1996;15(13):1421-34.
35. Gill J, Prasad V. Testing for blinding in sham-controlled studies for procedural interventions: the third-party video method. Cmaj. 2019;191(10):E272-e3.
36. Hrobjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. Int J Epidemiol. 2007;36(3):654-63.
37. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. BMJ. 2004;328(7437):432.
38. Canadian Cooperative Study Group. A randomized trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med. 1978;299(2):53-9.

39. Senn SJ. Turning a blind eye: Authors have blinkered view of blinding. BMJ. 2004;328(7448):1135-6.
40. Kolahi J, Bang H, Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. Community dentistry and oral epidemiology. 2009;37(6):477-84.
41. Cheon S, Park H-J, Chae Y, Lee H. Does different information disclosure on placebo control affect blinding and trial outcomes? A case study of participant information leaflets of randomized placebo-controlled trials of acupuncture. BMC Medical Research Methodology. 2018;18(1):13.
42. Heine M, Verschuren O, Hoogervorst EL, van Munster E, Hacking HG, Visser-Meily A, et al. Does aerobic training alleviate fatigue and improve societal participation in patients with multiple sclerosis? A randomized controlled trial. Multiple sclerosis (Houndmills, Basingstoke, England). 2017;23(11):1517-26.
43. Rees JR, Wade TJ, Levy DA, Colford JM, Jr., Hilton JF. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. Contemp Clin Trials. 2005;26(1):25-37.