Blinded interpretation of study results can feasibly and effectively diminish interpretation bias

Järvinen, Teppo L. N.

2014-07

Järvinen , T L N , Sihvonen , R , Bhandari , M , Sprague , S , Malmivaara , A , Paavola , M , Schuenemann , H J & Guyatt , G H 2014 , ' Blinded interpretation of study results can feasibly and effectively diminish interpretation bias ' , Journal of Clinical Epidemiology , vol. 67 , no. 7 , pp. 769-772 . https://doi.org/10.1016/j.jclinepi.2013.11.011

http://hdl.handle.net/10138/224256
https://doi.org/10.1016/j.jclinepi.2013.11.011

Journal of Clinical Epidemiology 67 (2014) 769–772

Blinded interpretation of study results can feasibly and effectively diminish interpretation bias

Teppo L.N. Järvinen[a,b], Raine Sihvonen[c], Mohit Bhandari[d,e], Sheila Sprague[d,e], Antti Malmivaara[f], Mika Paavola[a], Holger J. Schünemann[d,g], Gordon H. Guyatt[d,g,*]

[a]*Department of Orthopaedics and Traumatology, Helsinki University Central Hospital, Töölö Hospital, Topeliuksenkatu 5, 00260 Helsinki, Finland*
[b]*Department of Clinical Sciences, University of Helsinki, Tukholmankatu 8 B, PL 20 Biomedicum, Helsinki, Finland*
[c]*Department of Orthopaedics and Traumatology, Hatanpää City Hospital, Hatanpäänkatu 24, 33900 Tampere, Finland*
[d]*Department of Clinical Epidemiology & Biostatistics, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S4L8, Canada*
[e]*Division of Orthopaedic Surgery, Department of Surgery, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S4L8, Canada*
[f]*Centre for Health and Social Economics, National Institute for Health and Welfare, Mannerheimintie 166 I, 00271 Helsinki, Finland*
[g]*Department of Medicine, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S4L8, Canada*

Accepted 25 November 2013; Published online 20 February 2014

Abstract

Objective: Controversial and misleading interpretation of data from randomized trials is common. How to avoid misleading interpretation has received little attention. Herein, we describe two applications of an approach that involves blinded interpretation of the results by study investigators.

Study Design and Settings: The approach involves developing two interpretations of the results on the basis of a blinded review of the primary outcome data (experimental treatment A compared with control treatment B). One interpretation assumes that A is the experimental intervention and another assumes that A is the control. After agreeing that there will be no further changes, the investigators record their decisions and sign the resulting document. The randomization code is then broken, the correct interpretation chosen, and the manuscript finalized. Review of the document by an external authority before finalization can provide another safeguard against interpretation bias.

Results: We found the blinded preparation of a summary of data interpretation described in this article practical, efficient, and useful.

Conclusions: Blinded data interpretation may decrease the frequency of misleading data interpretation. Widespread adoption of blinded data interpretation would be greatly facilitated were it added to the minimum set of recommendations outlining proper conduct of randomized controlled trials (eg, the Consolidated Standards of Reporting Trials statement).

Keywords: Bias; Data interpretations; Double-blind method; Drug evaluation/methods; Randomized controlled trials as topic/methods; Research design

1. Introduction

Interpretation of data, a vitally important part of conducting research [1], is never totally objective and is therefore vulnerable to prior convictions, wishful thinking, and conflict of interest—in particular, the influence of commercial funding [2]. Presentations of results can be so profoundly misleading that the clinical message is the reverse of what should be conveyed [1,3–5]. One could argue that the best way to detect and correct such bias would be through peer-review process. The frequency of biased interpretation in the medical literature suggests, however, that many reviewers have the same sorts of biases as do the original researchers. Although guides for detecting bias and guides for consumers of research faced with misleading interpretations are available [4,6], it is often impossible to detect that the data analysis was flawed.

Author contributions: G.H.G. and T.L.N.J. conceptualized the study and led the writing of the manuscript; all authors contributed to the study. All authors have approved the final version of the manuscript. G.H.G. is the guarantor.

Competing interest statement: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no conflicts of interest related to this article.

Funding/Support and role of the sponsor: T.L.N.J.'s work is supported by the Sigrid Juselius Foundation and the Academy of Finland. The funding sources had no role in preparation, review, or approval of the manuscript. The authors' work was independent of the funders.

* Corresponding author. Department of Clinical Epidemiology & Biostatistics, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5. Tel.: +1-905-525-9140; fax: +1-905-524-3841.
E-mail address: guyatt@mcmaster.ca (G.H. Guyatt).

0895-4356
http://dx.doi.org/10.1016/j.jclinepi.2013.11.011

What is new?

- Although misleading interpretation of data (interpretative bias) was formally described more than 15 years ago in a seminal article by Gotzsche, currently there are few strategies for reducing the risk of interpretation bias.
- This article describes the application of an approach to execution of blinded interpretation of research data to safeguard against interpretation bias.
- The suggested procedure, best suited for the interpretation of data of randomized controlled trials, is simple, feasible, and efficient.

In this article, we describe a modification of a previously suggested approach to minimize the chance of misleading interpretation (interpretative bias) and describe its implementation.

1.1. Previous solutions

Gotzsche [3] first introduced the concept ''interpretive bias,'' although the specific term was introduced subsequently. He proposed that the authors of clinical trials should write two manuscripts, one assuming that treatment A is experimental and treatment B is control, and another article assuming the opposite (that treatment B is experimental and A is control). He suggested that both manuscripts be completed and approved by the authors before the randomization code is broken. Subsequently, Gotzsche [7] also went on to use this approach and, on three occasions, wrote two blinded manuscripts [8,9].

We implemented this approach in 2004 while in the process of preparing a manuscript that dealt with alternative approaches to eliciting patient utilities for health states [10]. The team statistician provided complete results labeled as group A and B; the rest of the research team was unaware of whether group A was exposed, or not exposed, to the marker states. One of us (H.J.S.) led us in producing many blinded draft versions, and finally two definitive manuscripts: One assuming that group A was exposed to marker states, the other that group B was exposed to marker states. We broke the code only after agreeing that there would be no further changes to the manuscripts, and submitted the appropriate manuscript. Although interesting and enlightening, we found the approach very onerous because it involved obtaining feedback from all coauthors on several revised, duplicate versions (groups A and B). In the many randomized trials our group had conducted subsequently, we have never repeated the process.

1.2. A more feasible alternative

Our next endeavor with blinded interpretation was in the reporting of the Study to Prospectively Evaluate Reamed Intramedually Nails in Patients With Tibial Fractures (SPRINT) trial [11,12], a multicenter randomized controlled trial (RCT) comparing the treatment of tibial shaft fractures with reamed or unreamed intramedullary nails. The writing committee of the trial was once again presented with an analysis of the results as treatment A and compared it with treatment B. Rather than writing two manuscripts, they discussed and came to agreement as to how they would interpret the results if treatment A proved to be reamed nailing and treatment B proved to be unreamed nailing. They recorded their decisions as ''Minutes of the Blinded Review of the Data'' document that was approved by all members of the Committee (see Appendix A at www.jclinepi.com). They then proceeded to break the randomization code, choosing the correct interpretation, and wrote the manuscript. The SPRINT Writing Committee members found this approach practical, feasible, and only marginally more time consuming than having a single interpretation.

The Finnish Degenerative Meniscal Lesion Study (FIDELITY) is a placebo–surgery controlled trial addressing the efficacy of arthroscopic partial meniscectomy (APM) in patients with degenerative meniscus lesion [13,14]. Prompted by the prior successful experience, one of the SPRINT authors (G.H.G.) proposed that the FIDELITY investigators consider using this approach in interpreting the data of the trial. As noted previously, the end result of the blinded interpretation process is a document we have called the ''Minutes of the Blinded Review of the Data'' (see Appendix B at www.jclinepi.com).

For the FIDELITY trial, the FIDELITY Writing Committee introduced two minor modifications to the procedure used in the SPRINT trial. First, they prepared a brief ''Background assumptions'' section and a succinct summary of the primary and secondary outcomes as well as key statistical analyses. These modifications were prompted by a belief that review of the theoretical basis of the trial would facilitate an objective and enlightened interpretation of the results. Second, to further increase the transparency and rigor of our blinded data interpretation, the FIDELITY Writing Committee introduced another safeguard to the process by asking an investigator not involved in any part of the FIDELITY trial (G.H.G.) to scrutinize our two interpretations (ie, to provide an ''external validation'').

This external validation (commentary) noted that for the primary outcome at 12 months, there was little issue: virtually no difference between groups, a conclusion that was secure whether A or B represented the group that received APM. The external reviewer suggested that the FIDELITY investigators may have preferred a definitive result of no benefit. Therefore, they were excessively inclined to dismiss findings at 2 months that suggested a difference in both Western Ontario Meniscal Evaluation Tool (WOMET) score

and pain after exercise. The external reviewer suggested that the FIDELITY investigators should acknowledge the possibility, indeed the likelihood, of transient benefit or harm after the procedure.

This suggestion by the external reviewer (a concern that the peer reviewers of the FIDELITY publication shared) prompted the FIDELITY investigators to reassess both the data analysis and interpretation of the findings of our trial. In making this reassessment, the FIDELITY investigators realized that they had neglected differences in the WOMET and pain after exercise scores between the two groups at baseline. An analysis adjusted for these baseline differences showed no transient 2-month APM-induced benefit in the WOMET and pain after exercise scores. This revision reinforced the initial interpretation of no difference between the two groups.

2. Discussion

Although guides to the reporting of randomized trials [15] and protocols for trials [16] are available, these initiatives focus on provision of accurate data, whereas investigators have paid relatively little attention—and provided even fewer suggestions for safeguards—to the risk of misleading data interpretation. We have developed and, in two randomized trials, implemented the blinded (treatments A and B) preparation of a summary of data interpretation assuming that treatment A is intervention and assuming treatment B is intervention. We have found the approach practical, efficient, and useful.

Our process is not the first of its kind. Pocock [17] described a similar process when he commented on the seminal article by Gotzsche [3] who described writing two complete manuscripts. Pocock suggested that, rather than writing two complete manuscripts, investigators should provide a detailed protocol of how they will interpret the results, record any additional analyses that will be undertaken whether experimental treatment is A or B, and then adhere to these plans once the code is broken. He suggested that one workable scenario would be for the primary outcomes part of the article's "Results" section to be drafted in two versions, together with "Summary" and "Conclusions" sections to match. Only then would the code be broken, secondary analyses be carried out, and consequent "fleshing out" of the manuscript be conducted. Pocock concluded his article by suggesting that this approach carries the prospect of having a manuscript made up of two parts: (1) primary results and conclusions ("beyond reproach" arising from blinded analysis and interpretation) and (2) secondary results and conjecture (prone to selectivity because of unblinded analysis and interpretation, and to be accompanied by a warning about potential bias).

Neither the approach suggested by Gotzsche [3] nor Pocock [17] has seen much use—indeed, apart from the three articles published by Gotzsche [7–9], we know of only one study in which either approach was implemented: our own experiment in writing two manuscripts. The approach we have described in this article provides a simplification to Pocock's proposal, further enhancing the efficiency of the procedure and its potential attractiveness and feasibility.

Another novel aspect of the approach described in this article is the external validation of the interpretation by an investigator not involved in any part of the trial. In the FIDELITY study, the external scrutiny proved important and helpful, calling the FIDELITY investigators' attention to a problem in interpretation that was ultimately resolved by a revised data analysis. Although in this case the external validation was carried out only after the first version of the paper was submitted for peer review, we recommend this procedure be performed before submission. One could go further and have the entire blinded interpretation done by a team of individuals separate from the investigators—as is done currently for data monitoring committees. Whether the additional safeguards against bias would be worth the added complexity and effort of such a procedure may warrant investigation.

What should be the next steps in reducing interpretative bias from medical literature? The focus currently is on obtaining bias-free data. For example, CONSORT states as follows (item 11a): "Unblinded data analysts may introduce bias through the choice of analytical strategies, such as the selection of favorable time points or outcomes, and by decisions to remove patients from the analyses." The CONSORT's intent is to provide standards for reporting, not conduct; thus, the guidance to conduct blinded analysis, although hard to misinterpret, remains implicit. A warning about unblinded interpretation, parallel to the warning about unblinded analysis, would be a worthwhile addition to CONSORT, and could have the same effect. Such a warning could come with a recommendation, again parallel to the current recommendation regarding blinded analysis, for an explicit statement about whether the authors conducted blinded interpretation. Such a statement would likely increase use of the method. An agreement among medical scientists that the details of a separate blinded interpretation plan (eg, as a part of statistical analysis and interpretation plan) are registered in a publicly available database (eg, ClinicalTrials.com) similar to the current practice regarding other details of RCTs could also promote increased use of blinded interpretation. Finally, post hoc changes (eg, adjusted analyses) made after the A/B code is broken need to be described as exploratory analyses.

Our proposal, although likely to reduce interpretation bias, is unlikely to represent a foolproof solution. Thus, even when authors have undertaken blinded interpretation, readers must remain alert to the possibility of interpretation bias.

Appendix

Supplementary material

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.11.011.

References

[1] Kaptchuk TJ. Effect of interpretive bias on research evidence. BMJ 2003;326:1453−5.

[2] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003;290:921−8.

[3] Gotzsche PC. Blinding during data analysis and writing of manuscripts. Control Clin Trials 1996;17:285−90. discussion 290−93.

[4] Montori VM, Jaeschke R, Schunemann HJ, Bhandari M, Brozek JL, Devereaux PJ, et al. Users' guide to detecting misleading claims in clinical research reports. BMJ 2004;329:1093−6.

[5] Montori VM, Isley WL, Guyatt GH. Waking up from the DREAM of preventing diabetes with drugs. BMJ 2007;334:882−4.

[6] Guyatt G, Rennie D, Meade MO, Cook DJ. Users' guides to the medical literature: a manual for evidence-based clinical practice. 2nd ed. New York, NY: McGraw-Hill; 2008.

[7] Gotzsche PC, Gjorup I, Bonnen H, Brahe NE, Becker U, Burcharth F. Somatostatin v placebo in bleeding oesophageal varices: randomised trial and meta-analysis. BMJ 1995;310:1495−8.

[8] Jorgensen AF, Mwakyusa D, Cegielski P, Gotzsche P, Hording M, Lallinger G, et al. The effect of fusidic acid on Tanzanian patients with AIDS. AIDS 1990;4:1037−8.

[9] Double blind dose-response study of zidovudine in AIDS and advanced HIV infection. Nordic Medical Research Councils' HIV Therapy Group. BMJ 1992;304:13−7.

[10] Schunemann HJ, Armstrong D, Degl'innocenti A, Wiklund I, Fallone CA, Tanser L, et al. A randomized multicenter trial to evaluate simple utility elicitation techniques in patients with gastroesophageal reflux disease. Med Care 2004;42:1132−42.

[11] Bhandari M, Guyatt G, Tornetta P 3rd, Schemitsch EH, Swiontkowski M, Sanders D, et al. Randomized trial of reamed and unreamed intramedullary nailing of tibial shaft fractures. J Bone Joint Surg Am 2008;90:2567−78.

[12] Bhandari M, Guyatt G, Tornetta P 3rd, Schemitsch E, Swiontkowski M, Sanders D, et al. Study to prospectively evaluate reamed intramedually nails in patients with tibial fractures (S.P.R.I.N.T.): study rationale and design. BMC Musculoskelet Disord 2008;9:91.

[13] Sihvonen R, Paavola M, Malmivaara A, Jarvinen TL. Finnish Degenerative Meniscal Lesion Study (FIDELITY): a protocol for a randomised, placebo surgery controlled trial on the efficacy of arthroscopic partial meniscectomy for patients with degenerative meniscus injury with a novel 'RCT within-a-cohort' study design. BMJ Open 2013;3(3).

[14] Sihvonen R, Paavola M, Malmivaara A, Itälä A, Joukainen A, Nurmi H, et al. Arthroscopic partial meniscectomy vs. sham for a degenerative meniscus tear. N Engl J Med 2013;369:2513−22.

[15] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869.

[16] Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013;346:e7586.

[17] Pocock SJ. Discussion. Control Clin Trials 1996;17:292−3.