

## LETTER TO THE EDITOR

## Use of a Spin Strategy Can Result in Unreliable Research Findings

Yasuyuki Okumura

Research Department, Institute for Health Economics and Policy, Association for Health Economics Research and Social Insurance and Welfare, Tokyo, Japan.

Correspondence: Yasuyuki Okumura, PhD, Research Department, Institute for Health Economics and Policy, Association for Health Economics Research and Social Insurance and Welfare, 11 Toyo Kaiji Bldg. 2F, 1-5-11 Nishishimbashi, Minato-ku, Tokyo 105-0003, Japan ([yokumura@blue.zero.jp](mailto:yokumura@blue.zero.jp)).

We read the article by Yamanouchi and colleagues with great interest and applaud the authors' efforts to assess the noninferiority of an antipsychotic tapering regimen compared with usual care in patients with schizophrenia who were treated with 2 or more antipsychotics (Yamanouchi et al., 2015). The results of the trial showed that there were no statistically significant differences in the co-primary outcomes (i.e., quality of life and severity of psychiatric symptoms) between the 2 groups despite the posthoc calculation with large statistical power and therefore supported the noninferiority hypothesis that a tapering regimen is not worse than usual care. Although the results of the trial had important implications for mental health policy in Japan, we believe that the results should be interpreted with caution due to several discrepancies between the study protocol submitted to their funding agency in April 2011 (Iwata, 2011) and the article published in March 2015 (Yamanouchi et al., 2015) of the same trial launched in November 2010 (Table 1). We used the study protocol (Iwata, 2011) instead of the trial registration document (Yamanouchi, 2010) to understand the original study design, because the trial registration document had inadequate information about the study design.

The first discrepancy was the stated use of quality of life, as assessed using the EuroQoL. In the study protocol, quality of life was specified as a primary outcome to test the superiority

hypothesis that a tapering regimen is superior to usual care. However, in the subsequent publication, the superiority hypothesis was changed to a noninferiority hypothesis without any notice.

The second discrepancy was the sample size calculation. In the study protocol, the sample size calculation for quality of life based on the use of a t test showed that a total sample size of 400 (200 in each group) would be required to provide 80% power at a 2-sided significance level of 5% to detect a standardized mean difference of 0.30, assuming a dropout rate of 10% (our sample size estimation of 392 was generated by using R version 3.2.2). In addition, the sample size calculation for symptom severity based on the use of the t test showed that a total sample size of 400 (200 in each group) would be required to provide 80% power at a 1-sided significance level of 2.5% to detect a mean difference of 0.8 (standard deviation of 5.8) with a noninferiority margin of 1.0, assuming a dropout rate of 10% (our sample size estimation of 363 was generated by using the TrialSize package in R).

However, in the subsequent publication, the required total sample size was reduced from 400 to 142, without an explanation for the protocol modifications. The assumptions for the sample size calculation in the published article were completely different from those in the study protocol. In the published article, the authors assumed the use of repeated measures ANOVA,

**Table 1.** Three Discrepancies between the Study Protocol Submitted to Their Funding Agency in April 2011 and the Article Published in March 2015 of the Same Trial Launched in November 2010

Concern	Original study design	Published study design
Quality of life	Primary outcome to test a superiority hypothesis	Co-primary outcome to test a noninferiority hypothesis
Sample size calculation	400	142
Severity of psychiatric symptoms	Secondary outcome to test a noninferiority hypothesis	Co-primary outcome to test a noninferiority hypothesis

Received: January 31, 2017; Revised: February 7, 2017; Accepted: February 9, 2017

© The Author 2017. Published by Oxford University Press on behalf of CINP.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

a power of 80% at a significance level of 2.5% to detect an effect size of 0.20 (called Cohen's  $f$ ) without specification of a noninferiority margin, the number of repeated measurement to be 6, a correlation among repeated measures of 0.50, and a nonsphericity correction of 1. Although the authors asserted that the sample size calculation was for testing the noninferiority hypothesis, we believe that the sample size calculation was for testing the superiority hypothesis.

Now, we consider the statistical power with the reduced sample size. When assuming that all parameters except the statistical power and dropout rate are set to the same values as those in the study protocol (i.e., sample size of 142 and dropout rate of 0%), the statistical power is reduced from 80% to 39% for quality of life and to 42% for symptom severity. To calculate more conservative estimates, when assuming the sample size is set at the actual enrolled number of 163 participants, statistical power is reduced from 80% to 44% for quality of life and to 47% for symptom severity.

The third discrepancy was the importance of the severity of psychiatric symptoms, as assessed using the Manchester scale. In the study protocol, symptom severity was specified as a secondary outcome to test the noninferiority hypothesis. However, in the subsequent publication, symptom severity was insidiously upgraded to a co-primary outcome.

In conclusion, we believe that the trial was a "negative trial," with misleading reporting practices that are not negligible. In fact, the results of the trial do not support the noninferiority hypothesis in the published article due to the modified hypothesis, inadequate statistical power, and selective outcome reporting. Such a misleading reporting practice is called "spin," which is defined as specific reporting strategies that could distort the interpretation of results and mislead readers, for whatever motive (Boutron et al., 2010; Hernandez et al., 2013). One possible reason for the use of the spin strategy is that the research team consisted entirely of clinicians (i.e., 6 psychiatrists and a pharmacist) without epidemiologist or bio-statistician team members. In general, researchers with inadequate knowledge of epidemiology and/or biostatistics could be tempted to unintentionally use a spin strategy when their results have negative findings. However, unreliable results can be a consequence of the use of a spin strategy, and the editor could decide to retract the article (Cassani et al., 2015, 2016; Dimova and Allison, 2016). The retraction guideline developed by the Committee on Publication Ethics recommends that journal editors should consider retracting a publication if they have clear evidence that the findings are unreliable, either as a result of misconduct or honest error (Committee on Publication Ethics, 2009). A trial for which the superiority hypothesis is changed to a noninferiority hypothesis is typically vulnerable to unreliable findings; therefore, editors should consider retracting the publication in such situations. Researchers, reviewers, and editors should be aware of the need to ensure that the reported results are reliable and their conclusions are an appropriate reflection of the design and results.

## Statement of Interest

During the past 3 years, Y.O. has served as a chairperson of the Reporting Quality Initiative of Researchers in Clinical Epidemiology; an associate editor of the *Journal of Epidemiology*, the *Japanese Journal of Behavior Therapy*, and the *Japanese Journal of Cognitive Therapy*; and an adviser for reviewers of the *Journal of Japan Academy of Gerontological Nursing*. He received personal fees from Janssen Pharmaceuticals, Inc., the Medical Technology Association, and Cando Inc. He has received research grants from the Japan Agency for Medical Research and Development; Ministry of Health, Labour and Welfare; Japan Society for the Promotion of Science; Institute for Health Economics and Policy; and Mental Health and Morita Therapy.

## References

- Boutron I, Dutton S, Ravaud P, Altman DG (2010) Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 303:2058–2064.
- Cassani RS, Fassini PG, Silvah JH, Lima CM, Marchini JS (2015) Impact of weight loss diet associated with flaxseed on inflammatory markers in men with cardiovascular risk factors: a clinical study. *Nutr J* 14:5.
- Cassani RS, Fassini PG, Silvah JH, Lima CM, Marchini JS (2016) Retraction note: impact of weight loss diet associated with flaxseed on inflammatory markers in men with cardiovascular risk factors: a clinical study. *Nutr J* 15:59.
- Committee on Publication Ethics (2009) Guidelines for retracting articles. Available at <http://publicationethics.org/files/retraction%20guidelines.pdf>. Accessed December 25, 2016.
- Dimova RB, Allison DB (2016) Inappropriate statistical method in a parallel-group randomized controlled trial results in unsubstantiated conclusions. *Nutr J* 15:58.
- Hernandez AV, Pasupuleti V, Deshpande A, Thota P, Collins JA, Vidal JE (2013) Deficient reporting and interpretation of noninferiority randomized clinical trials in HIV patients: a systematic review. *PLoS One* 8:e63272.
- Iwata N (2011) The clinical study to correct multiple and large amount of administering to antipsychotic safely and effectively (in Japanese). Available at <http://mhlw-grants.niph.go.jp/niph/search/NIDD00.do?resrchNum=201027094A>. Accessed December 25, 2016.
- Yamanouchi Y (2010) The clinical study to correct multiple and large amount of administering to antipsychotic safely and effectively. Available at [https://upload.umin.ac.jp/cgi-open-bin/ctr\\_e/ctr\\_view.cgi?recptno=R000005391](https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000005391). Accessed December 25, 2016.
- Yamanouchi Y, Sukegawa T, Inagaki A, Inada T, Yoshio T, Yoshimura R, Iwata N (2015) Evaluation of the individual safe correction of antipsychotic agent polypharmacy in Japanese patients with chronic schizophrenia: validation of safe corrections for antipsychotic polypharmacy and the high-dose method. *Int J Neuropsychopharmacol* 18:1–8.