Between-group difference in mean values or changes in pain intensity? Evaluating the distribution of change from baseline in a neuropathic cancer pain clinical trial

Hiromichi Matsuoka[1,2,3,4], Tatsuya Morita[5], Shunsuke Oyamada[6], Takuhiro Yamaguchi[7], Atsuko Koyama[1,2]

[1]Department of Psychosomatic Medicine, Faculty of Medicine, Kindai University, Osaka, Japan; [2]Palliative Care Center, Faculty of Medicine, Kindai University, Osaka, Japan; [3]National Cancer Center Hospital, Tokyo, Japan; [4]Improving Palliative, Aged and Chronic Care through Clinical Research and Translation (IMPACCT), Faculty of Health, University of Technology Sydney, Ultimo, NSW, Australia; [5]Palliative and Supportive Care Division, Seirei Mikatahara General Hospital, Hamamatsu, Japan; [6]Japanese Organisation for Research and Treatment of Cancer (JORTC), NPO, Tokyo, Japan; [7]Division of Biostatistics, Graduate School of Medicine, Tohoku University, Sendai, Japan

Correspondence to: Hiromichi Matsuoka, MD, PhD. 377-2 Ohno-higashi, Osakasayama City, Osaka, 589-8511, Japan. Email: hiromima@ncc.go.jp.

Comment on: Mbowe OB, Gewandter JS, Turk DC, *et al*. Are there really only 2 kinds of people in the world? Evaluating the distribution of change from baseline in pain clinical trials. Pain 2020;161:195-201.

When evaluating pain relief in pain research, both between-group differences in mean values and changes in pain intensity (absolute or relative values) are recommended (1). However, there is no consensus on which one is superior, thereby causing a few challenges such as difficulty in comparing multiple research views and differences in sample size calculations, resulting in different results within a single study.

Our recent studies (2,3) have also yielded different results within a single study, making interpretation difficult. One such study of ours, the DIRECT study, which was a multicenter double-blinded randomized placebo-controlled two-parallel group trial, investigated the efficacy of duloxetine for cancer-related neuropathic pain (CNP) nonresponsive or intolerant to opioid-pregabalin combination therapy. In that study, duloxetine was commenced at a dose of 20 mg once daily. On day-3, the degree of pain relief was evaluated using the Pain Relief Scale (4). Patients who reported complete or substantial relief continued to receive the same dose while the dose was increased to 40 mg for the remaining patients. The primary endpoint was a comparison of pain intensity [Brief Pain Inventory (BPI)-item 5] (5) at day-10 in both groups. Seventy eligible patients were enrolled and randomized, and 65 participants completed the study. Among the 70 patients, 67 were evaluable: 34 in group duloxetine (Group

D) and 33 in group placebo (Group P). There were no primary endpoint data for one patient in each group. Of the 67 evaluable subjects, 32 patients (94.1%) in Group D and 29 patients (87.9%) in Group P completed the protocol treatment on day-10. Twenty-two patients (63%) of patients in Group D and 23 patients (66%) of those in Group P increased the dose to 40 mg per day. Three different types of analyses were applied revealing the following results.

Firstly, in the complete case (CC) analysis, BPI-item 5 on day-10 was 4.03 [90% confidence interval (CI): 3.33–4.74] for Group D and 4.88 (90% CI: 4.37–5.38) for Group P (P=0.053). Secondly, in the sensitivity analysis using baseline observation carried forward (BOCF) (6), average pain scores on day-10 were 4.06 (90% CI: 3.37–4.74) for Group D and 4.91 (90% CI: 4.41–5.41) for Group P (P=0.04). BOCF is based on the assumption that the pain returns to the baseline state if treatment is stopped because neuropathic cancer pain is unlikely to improve in the natural course during this short period. Hence, we imputed missing data on day-10 with the score from day-0. However, responder analysis, comparing the percentages of patients with clinically meaningful pain reduction (≥30% and ≥50%), showed different results. On the tenth day, 15 patients (44.1%) in Group D and 6 patients (18.2%) in Group P reported pain improvement ≥30% (P=0.02) whereas 11 patients (32.4%) in Group D and 1 patient (3%) in Group P

P reported pain improvement ≥50% (P=0.002).

What is the clinical importance of patient improvements? Multiple anchor-based analyses (7) have confirmed that patients consider pain intensity reductions on numerical or visual analogue scales of at least 2 points or 30% to be moderately clinically meaningful, and that a reduction of 1 point or 10–20% represents a minimally important change (8). Hui *et al.* (9) also reported that the optimal cut-off was ≥1 point for improvement in cancer pain. Therefore, setting a standardized endpoint for pain intensity reductions is essential.

Pain patients may be placed into two categories: good responders and poor responders to effective pain treatment, with a few patients in between. This hypothesis is based on research that claims that changes in pain intensity have a bimodal distribution (10-13). Mbowe *et al.* (14) recently published a study entitled "Are there really only 2 kinds of people in the world? Evaluating the distribution of change from baseline in pain clinical trials". This study attempted to critically examine the claim of bimodality of the distribution of change in pain intensity from 4 randomized trials of pregabalin and duloxetine for chronic musculoskeletal and neuropathic pain conditions (15-18). After imputation of missing data using the BOCF method, the distributions of the percentage change from baseline in the active treatment group were summarized using histograms that were constructed using unequal-width bins (<15%, 15–29%, 30–49%, and ≥50% reduction in pain intensity from baseline) (10-13) and equal-width bins (20%) which is required for proper representation of distributions using histograms. These unequal-width bins seem to be selected to correspond to the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommended definitions of a substantial reduction in pain (≥50% reduction in pain intensity from baseline), a moderate reduction in pain (≥30% reduction in pain intensity from baseline), and a minimal reduction in pain (10–20%–reduction in pain intensity from baseline) (8).

Given the shortcomings in evaluation of percentage change from baseline as an indicator, Mbowe *et al.* constructed histograms using equal-width bins for the absolute change from baseline for each trial, after imputation of missing data, which illustrated that with a certain choice of cut-points, it was possible to make the distribution of absolute change from baseline in pain intensity look strongly bimodal, even when the true distribution seemed to be unimodal. The authors mentioned that the incorrect construction of histograms

using unequal-width bins was found to be the principal flaw leading to the bimodality claim, along with the use of the oft-criticized BOCF method for imputing missing data also serving as a contributing factor. Correctly constructed histograms of absolute change in pain intensity using equal-width bins combined with more principled methods for handling missing data resulted in distributions that had a more unimodal appearance. They concluded that the evidence presented in their study neither supported nor refuted the hypothesis that distinct populations of "responders" and "non-responders" to pain interventions existed (14). However, their findings demonstrated that the evidence they presented in their earlier work (10-13) to support this hypothesis was rather misleading.

When we evaluated our research findings (2) in a similar manner, we observed that there was a bimodal distribution when using unequal-bin width and a unimodal distribution when using equal-bin width in the percentage change (relative model) and the absolute change (absolute model) with a certain choice of cut-points under the recommendation of IMMPACT for minimal clinically important difference (MCID) for chronic pain research (1) (absolute change; <1, 1.0, 2.0, and ≥3.0) (*Figure 1*). With regard to bimodal distribution, our study results compared favorably to those of Mbowe *et al.* (5). They however felt that their results would have been more informative had it been possible to analyze other trials of additional treatments. Similarly, a larger sample size for this study (i.e., neuropathic cancer pain) would have revealed more informative results.

In our study, data was missing in each group, but the number of participants with missing endpoint data was extremely low in both groups: 1 patient in each group (approximately 3%). In such a situation, we agreed that by eliminating the cases where missing data occurred, it would not have had a significant effect on the results of the analysis. We therefore adopted the CC as the primary analysis. In addition, since neuropathic cancer pain, which is the subject of our study, is unlikely to improve in the natural course of the disease, we deduced that it matched the premise of the BOCF method i.e., "return to the baseline state after treatment", and therefore adopted it as a sensitivity analysis to investigate the stability of the result. However, a certain improvement was observed from the baseline to day-10 even in the placebo group. Therefore, "placebo effects and the power of expectation, and the increased attention that trial participants receive from clinic staff, among others" as described by Mbowe *et al.* may be
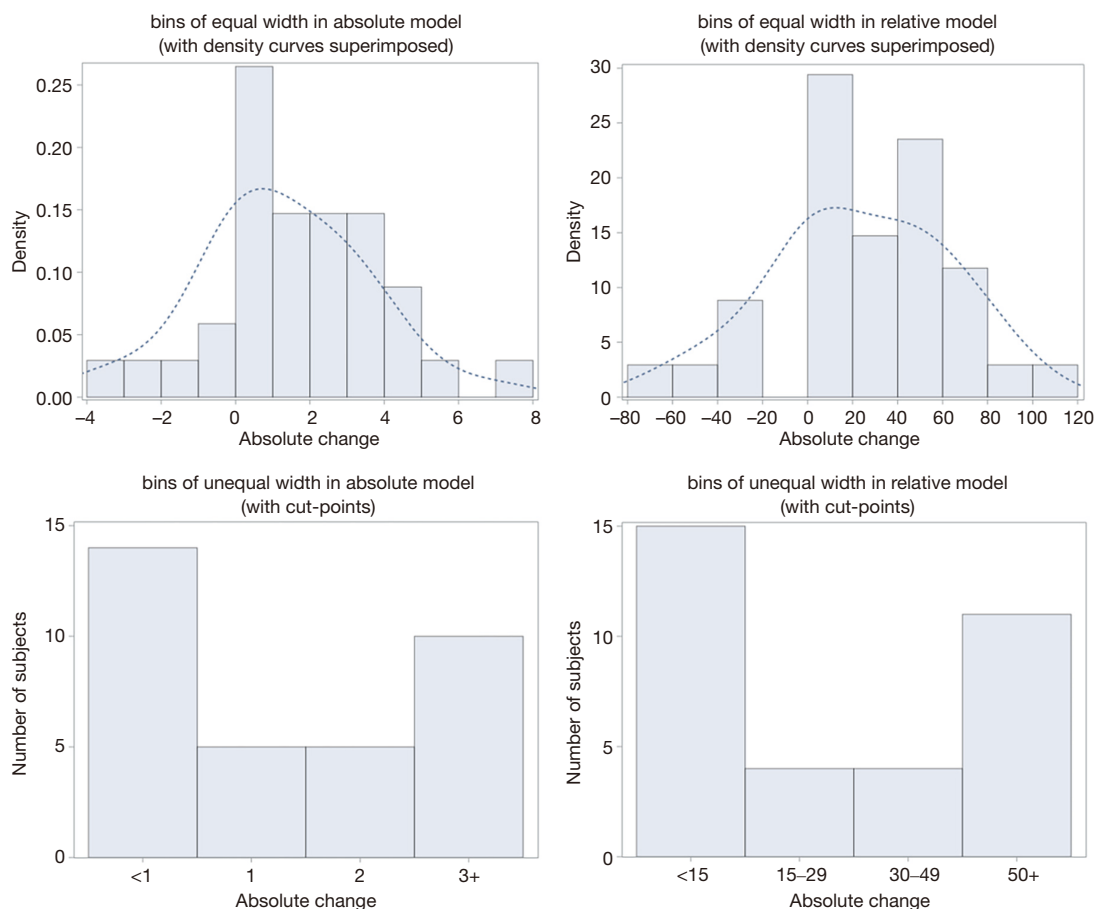
Figure 1 Histograms of absolute and relative change from baseline in pain intensity (with positive values indicating improvement) after BOCF imputation for missing data in duloxetine treated subjects for neuropathic cancer pain. The histograms on the top row, with density curves superimposed, were constructed using bins of equal width, and the "bimodalized" histograms on the bottom row were constructed using bins of unequal width.

possible, and the assumption of "returning to the baseline state after the end of treatment" may not hold.

The BOCF method has poor statistical properties, for example, underestimating the variance of the treatment effect, and possibly leading to an inflation of the type I error probability. Under the Missing at Random (MAR) assumption, there is a probability of leading to biased results. However, the number of participants with missing endpoint data in our study was very low, and it was assumed that even if the use of BOCF lead to biased results, its effect would not be significant. Therefore, we concluded that a simple method considering the characteristics of the disease would be better than emphasizing the assumed missing data mechanism that cannot be verified from the data. Furthermore, in our study, BOCF was not used as a primary analysis tool, but as a tool for sensitivity analysis, and the

results were not used to determine the outcome of the study.

Although our study showed a difference in the between-group difference in mean values using BOCF (P=0.048) and changes in pain intensity (absolute or relative values), it failed to show a difference in the between-group difference in mean values using CC (P=0.053). Group comparison using mean value has a high statistical power because the entire information in the data can be used, and it allows easier comparison with other studies than group comparison using proportion based on clinically significant threshold categorization. On the other hand, group comparison using proportion is useful when categorization is clinically important and makes interpretation of the result easier. For individual patients, it is more important that they assess the effectiveness of the treatment i.e., does it reduce pain intensity effectively? Therefore, we believe it is important
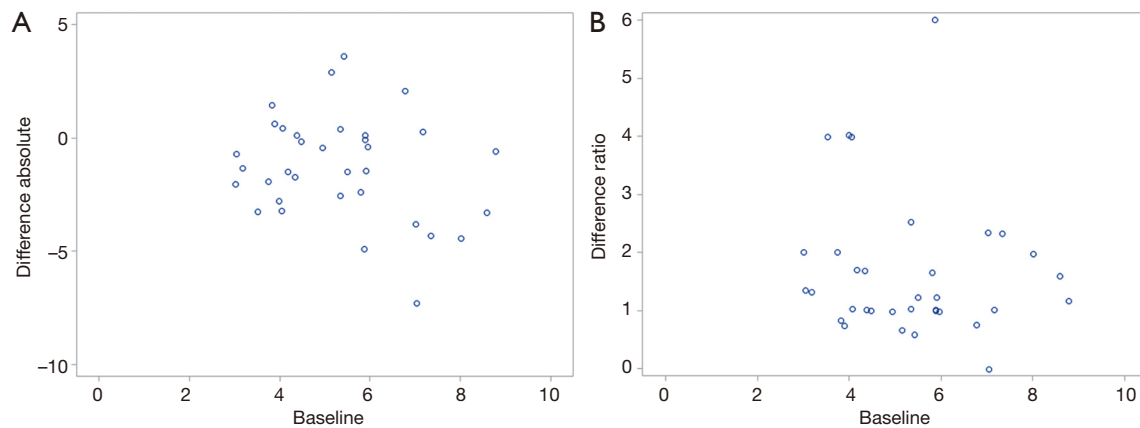
Figure 2 Scatter plot of the difference (day-10 pain intensity minus baseline measures) against baseline intensity. (A) the variability of difference absolute seems to be stable and (B) the variability of difference ratio tends to be wide partially, but seems to be nearly stable.

that patients' assessment of pain relief scores, with or without a bimodal distribution, be analyzed not only by population mean, but by changes in pain intensity.

Regarding which criterion (absolute or relative) should be used to determine an individual's pain intensity response, in the study by Johnson *et al.* (19), absolute values, not relative values, showed uniform variation in all baseline intensities. Therefore, absolute values are more recommended for minimal clinically important difference (MCID). In the pain research field, neither is recommended. We have tried to illustrate which criterion (absolute or relative) would be more suitable to determine an individual's pain intensity response however a clear difference was not observed between the two (*Figure 2*).

Incidentally, in the field of palliative care, there is also the concept of "Personalized Pain Goal" (20). It is a tool used to tailor pain management to individual needs. Participants are asked to describe on a 0–10 scale, the level/intensity of pain that will allow the patient to achieve comfort in physical, functional, and psychosocial domains. Zero represents no pain and 10 represents worst pain. Some argued that neither between-group difference in mean values nor changes in pain intensity (e.g., absolute or relative values) correctly evaluated the patient's discomfort. Furthermore, another concern is how to remove the effects of rescue medications (prohibiting or otherwise using composite scale).

In conclusion, we agree with Mbowe *et al.* that changes in pain intensity have a bimodal distribution because of the difference in bin width. However, we believe it is important that patients' evaluations, with or without a bimodal distribution, be assessed not only by population mean but by changes in pain intensity. There is no consensus between

using absolute values or relative percentages in our JORTC-PAL08 (DIRECT) study. Although there is only one subject in the BOCF analysis, our results are still bimodal. Further studies are needed to achieve greater clarity on these points.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Annals of Palliative Medicine*. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/apm-20-930). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

References

1. Dworkin RH, Turk DC, McDermott MP, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009;146:238-44.

2. Matsuoka H, Iwase S, Miyaji T, et al. Additive Duloxetine for Cancer-Related Neuropathic Pain Nonresponsive or Intolerant to Opioid-Pregabalin Therapy: A Randomized Controlled Trial (JORTC-PAL08). J Pain Symptom Manage 2019;58:645-53.

3. Matsuoka H, Iwase S, Miyaji T, et al. Predictors of duloxetine response in patients with neuropathic cancer pain: a secondary analysis of a randomized controlled trial-JORTC-PAL08 (DIRECT) study. Support Care Cancer 2020;28:2931-9.

4. Matsuoka H, Ishiki H, Iwase S, et al. Study protocol for a multi-institutional, randomised, double-blinded, placebo-controlled phase III trial investigating additive efficacy of duloxetine for neuropathic cancer pain refractory to opioids and gabapentinoids: the DIRECT study. BMJ Open 2017;7:e017280.

5. Committee for Medicinal Products for Human Use, European Medicines Agency. Guideline on missing data in confirmatory clinical trials. European Medicines Agency, London, 2010.

6. Cleeland CS. Measurement of pain by subjective report. In: Chapman CRL, Loeser. JD (eds). Advances in Pain Research and Therapy. Volume 12: Issues in Pain Measurement. Raven Press, New York, 1989; 391-403.

7. de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 2006;4:54.

8. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105-21.

9. Hui D, Shamieh O, Paiva CE, et al. Minimal clinically important differences in the Edmonton Symptom Assessment Scale in cancer patients: a prospective, multicenter study. Cancer 2015;121:3027-35.

10. Moore RA, Smugar SS, Wang H, et al. Numbers needed-to-treat analyses—do timing, dropouts, and outcome matter? Pooled analysis of two randomized, placebo-controlled chronic low back pain trials. Pain 2010;151:592-7.

11. Moore RA. What works for whom? determining the efficacy and harm of treatments for pain. Pain 2013;154:S77-86.

12. Moore RA, Cai N, Skljarevski V, et al. Duloxetine use in chronic painful conditions–individual patient data responder analysis. Eur J Pain 2014;18:67-75.

13. Moore RA, Derry S, Simon LS, et al. Nonsteroidal anti-inflammatory drugs, gastroprotection, and benefit risk. Pain Pract 2014;14:378-95.

14. Mbowe OB, Gewandter JS, Turk DC, et al. Are there really only 2 kinds of people in the world? Evaluating the distribution of change from baseline in pain clinical trials. Pain. 2020;161:195-201.

15. Skljarevski V, Ossanna M, Liu-Seifert H, et al. A double-blind, randomized trial of duloxetine versus placebo in the management of chronic low back pain. Eur J Neurol 2009;16:1041-8.

16. Chappell AS, Ossanna MJ, Liu-Seifert H, et al. Duloxetine, a centrally acting analgesic, in the treatment of patients with osteoarthritis knee pain: a 13-week, randomized, placebo-controlled trial. Pain 2009;146:253-60.

17. Richter RW, Portenoy R, Sharma U, et al. Relief of painful diabetic peripheral neuropathy with pregabalin: a randomized, placebo-controlled trial. J Pain 2005;6:253-60.

18. van Seventer R, Feister HA, Young JP Jr, et al. Efficacy and tolerability of twice-daily pregabalin for treating pain and related sleep interference in postherpetic neuralgia: a 13-week, randomized trial. Curr Med Res Opin 2006;22:375-84.

19. Johnson MJ, Bland JM, Oxberry SG, et al. Measuring improvement in dyspnoea: should absolute or relative values be used? Eur Respir J 2014;44:1700-3.

20. Dalal S, Hui D, Nguyen L, et al. Achievement of personalized pain goal in cancer patients referred to a supportive care clinic at a comprehensive cancer center. Cancer 2012;118:3869-77.