JAMA Guide to Statistics and Methods

Treatment Effects in Multicenter Randomized Clinical Trials

Stephen J. Senn, PhD; Roger J. Lewis, MD, PhD

It is common for treatments to be evaluated in clinical trials that involve many sites or centers, primarily because one center rarely can enroll sufficient numbers of patients to complete the trial.[1] The use of multiple clinical sites introduces complexity because outcomes at different sites may be systematically different, eg, due to differences in patient populations, ancillary treatment practices, or other factors. Thus, appropriate statistical analyses of multicenter clinical trials consider these center effects to yield a better understanding of the overall mean treatment effect and the variability in treatment effects and patient outcomes among sites.[1]

In the May 15, 2018, issue of *JAMA*, Dodick et al[2] published the results of a clinical trial that compared migraine prevention by 2 different dosing regimens of fremanezumab vs placebo. The number of migraine days were recorded during a 28-day baseline period and a 3-month treatment period. The primary outcome for the study was the change from baseline in the mean number of monthly migraine days during treatment. The 875 participating patients were recruited from 123 centers in 9 countries. Using a primary analysis that accounted for each patient's mean number of migraines during the baseline period,[2,3] treatment, country (US vs non-US), and other factors, the authors reported a difference with monthly dosing vs placebo of −1.5 days (95% CI, −2.01 to −0.93 days; $P < .001$) and with single higher dosing vs placebo of −1.3 days (95% CI, −1.79 to −0.72 days; $P < .001$). They also conducted a post hoc sensitivity analysis that accounted for effects of the specific country of enrollment.[2]

Estimating Treatment Effects in Multicenter Clinical Trials

Why Are Differences Between Centers Considered When Estimating Treatment Effects?

The goals of the statistical analysis of a multicenter clinical trial include providing a valid estimate of the treatment effect (ie, the mean difference in outcomes between patients treated in the 2 groups) and understanding and quantifying the remaining uncertainty or precision in the estimated treatment effect.[1] Patients treated at different centers may differ in their overall prognoses but experience the same relative benefit of a treatment compared with standard care. Alternatively, patients at different centers may differ in both their overall prognoses and in the treatment effect. Only the first case is considered in this article.

The randomization of patients to treatments in multicenter trials is usually stratified by center to achieve balance in the numbers of patients receiving each treatment within each center and, in what follows, it is assumed this has been done.[4] Balance improves the statistical efficiency of the trial, increasing precision in the estimation of treatment effects given a particular sample size. It also reduces the risk in modestly sized trials that chance imbalance in treatment allocation at centers with smaller numbers of patients results in a bias if that center has better or worse outcomes on average than other centers.[1,4]

Center effects, including systematic differences in outcomes between patients enrolled in different centers, have the potential to affect the estimate of the treatment effect (eg, if patients vary in severity of illness from center to center, any imbalance in treatment allocation within centers could affect the estimate of the treatment effect). However, a potentially more important and less well-appreciated effect of differences between centers is on the uncertainty or the precision in the estimate of the treatment effect. Even when stratified randomization is used successfully to achieve balance between groups across centers,[4] the resulting confidence intervals (CIs) and *P* values can change substantially depending on whether the center effect is included in the statistical model used to estimate the treatment effect, potentially affecting the overall interpretation of the clinical trial result.[1]

The uncertainty in the estimate of the treatment effect is defined by the variability in the estimates that would be obtained, hypothetically, if equivalent multicenter clinical trials were independently repeated many times. There are two distinct ways that such repetition might be conducted. First, the same centers might be used but the random treatment allocation (eg, which treatment is assigned to each patient in sequence) would be changed from repetition to repetition, maintaining stratification to balance treatments within center each time. Second, different centers could also be used for each trial, selected from a larger pool of centers with similar characteristics. Only the first approach is considered here, which addresses the uncertainty in the treatment effect obtained from the original trial with those particular centers. The second type of repetition, with variation in the participating centers, would address a more general type of uncertainty—including uncertainty due to difference in the treatment effect among centers—that is beyond the scope of this discussion.

Suppose the same multicenter clinical trial results were analyzed using 2 different statistical models. First, a simple model is used that does not include a term for center effect, so all variability is assumed to arise only from the inherent variation among patients but without any systematic differences from center to center. Second, a more appropriate model is used that separates the variability arising from differences among patients within centers from the variability arising from differences among centers. The uncertainty in the treatment effect obtained from the 2 models, representing the variability in the mean treatment effect that would be seen in the repeated equivalent trials, and reflected in the width of the 95% CI around the estimated treatment effect, will be different. The first model, without the center effect, will overestimate the variability in the estimated mean treatment effect because it assumes that all variability is an inherent characteristic of the patient population. This will result in an overly wide CI and decreased statistical power (the overly wide CI will be more likely to include a zero or null treatment effect, equivalent to a nonstatistically significant *P* value). In contrast, the model incorporating the center effect will correctly recognize that some of the variability is a characteristic of patients within each center and some is a characteristic of differences in populations enrolled at different centers, and eliminate the latter. This will

© 2019 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a American University of Beirut User on 03/27/2019

reduce the estimated variability for patients within each center, resulting in a narrower, more accurate CI and increased power.

How Are Center Effects Incorporated into Estimates of Treatment Effects?

Various statistical models can be used to account for center effects when estimating the treatment effect. The simplest way is to consider centers as fixed effects—each center is associated with its own effect on patient outcomes—in an appropriate model. This approach can be use in linear models, survival analysis, or logistic regression for dichotomous outcomes and, in each case, allows for the greater similarity among treatment outcomes within a center compared with across centers.[5]

Limitations of Estimates of Treatment Effects From Multicenter Clinical Trials

It is often stated that the purpose of enrolling patients at multiple centers is to increase the external generalizability of the trial results; however, centers are typically selected for factors (eg, academic affiliations, large patient volumes) intended to speed enrollment and that limits the generalizability of results to other clinical settings. Thus, even estimates from multicenter clinical trials may lack external validity when applied to qualitatively different practice settings.

Each of the models used to adjust for center effects have underlying assumptions regarding how the differences between centers affect patient outcomes or, similarly, how patients within centers are more similar on average than patients between centers. If these assumptions do not hold true, then the results of the analyses may be biased or estimates of uncertainty may be incorrect, affecting the statistical significance of results or the width of CIs.

How Were the Multicenter Data Analyzed in the Study by Dodick et al?

The 875 participating patients were recruited in 123 centers in 9 countries and randomized in a 1:1:1 ratio to the 3 treatments.[2] Patient randomization was stratified by sex, country, and baseline preventive medication used. The primary analysis used analysis of covariance[3] to adjust for the influence of the 3 stratification factors, the number of episodes occurring in the baseline period, and the number of years since first onset of migraine. The effect of country was simplified by reducing this variable to only 2 groups: US vs non-US. To more completely evaluate the potential "country effect" (analogous to the "center effect" described above), the authors conducted a "post hoc sensitivity analysis using a mixed-effects model that included country instead of region as a random effect."[2] The results of this analysis are presented in eTable 4 in the article's Supplement 3, demonstrating a difference from placebo with monthly dosing of −1.5 (95% CI, −2.00 to −0.93) and a difference with the single higher dose of −1.3 (95% CI, −1.79 to −0.72).[2] These results are almost identical to those of the primary analysis, suggesting that either there was little effect of country on the results or that all effect of the country was captured by simply distinguishing US sites from non-US sites.

How Should the Results From This Study Be Interpreted?

The treatment effects estimated by Dodick et al,[2] and the associated CIs, were nearly identical regardless of whether the effect of location of enrollment was dichotomized as within or outside of the US, or captured as the country of enrollment, with 9 possibilities. The consistency suggests that the variability in outcomes associated with the location of enrollment is either small or is captured similarly by both models. The results from the model adjusting for the country of enrollment is the preferred estimate, because there is no harm in the adjustment if the country-to-country variability is unimportant and the adjustment is critically important for correctly determining the uncertainty in the estimate of the mean treatment effect if the country-to-country variability turns out to be substantial. In both the primary and post hoc sensitivity analyses, the models partitioned the variability in patient outcomes between that associated with the location of enrollment (US vs non-US or by country) and that associated with differences between patients, resulting in more accurate CIs than would have obtained had these effects not been included.

ARTICLE INFORMATION

Author Affiliations: Luxembourg Institute of Health, Strassen, Luxembourg (Senn); Medical Statistics Group, ScHARR, The University of Sheffield, Sheffield, United Kingdom (Senn); Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California (Lewis); Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California (Lewis); Berry Consultants LLC, Austin, Texas (Lewis).

Corresponding Author: Stephen J. Senn, PhD, 29 Merchiston Crescent, Edinburgh, E10 5AJ, United Kingdom (stephen@senns.demon.co.uk).

Section Editor: Edward H. Livingston, MD, Deputy Editor, *JAMA*.

Published Online: March 1, 2019. doi:10.1001/jama.2019.1480

Conflict of Interest Disclosures: Dr Senn reported serving as a consultant to the pharmaceutical industry.

REFERENCES

1. Senn S. Some controversies in planning and analysing multi-centre trials. *Stat Med*. 1998; 17(15-16):1753-1765. doi:10.1002/(SICI)1097-0258(19980815/30)17:15/16<1753::AID-SIM977>3.0.CO;2-X

2. Dodick DW, Silberstein SD, Bigal ME, et al. Effect of fremanezumab compared with placebo for prevention of episodic migraine: a randomized clinical trial. *JAMA*. 2018;319(19):1999-2008. doi:10.1001/jama.2018.4853

3. Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ*. 2001;323(7321):1123-1124. doi:10.1136/bmj.323.7321.1123

4. Broglio K. Randomization in clinical trials: permuted blocks and stratification. *JAMA*. 2018;319 (21):2223-2224. doi:10.1001/jama.2018.6360

5. Meurer WJ, Lewis RJ. Cluster randomized trials: evaluating treatments applied to groups. *JAMA*. 2015;313(20):2068-2069. doi:10.1001/jama.2015.5199

© 2019 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a American University of Beirut User on 03/27/2019