JOURNAL OF CME
2023, VOL. 12, 2236893
https://doi.org/10.1080/28338073.2023.2236893

EUROPEAN CME Forum
Taylor & Francis
Taylor & Francis Group

LETTER TO THE EDITOR

OPEN ACCESS Check for updates

Continuing Medical Education Outcomes are Much More Than Statistical Significance

Katie Stringer Lucero [a] and Donald E. Moore[a,b]

[a]Medscape, LLC, New York, NY, USA; [b]Medical Education and Administration, Emeritus, Vanderbilt University School of Medicine

Dear Editor,

An article published by the *Journal of CME* in 2023 by Robles and colleagues [1] reports the results of an evaluation of a continuing education programme, a free live continuing education (CE) series of activities primarily for primary care advanced practice providers offered in 2019 by a medical education company (Practicing Clinicians Exchange) and discusses the potential value of pooled samples in comparison to paired samples in examination of percentage of correct responses pre- and post-continuing medical education (CME). The focus of the article is statistical significance and sample size. It is well known in the research literature that statistical significance alone is not sufficient to accept that a certain set of outcomes was the result of participation in a series of educational activities. Rather, it is now regarded as necessary to report effect size as well. The effect size is the magnitude of the difference between two groups, like the pre- and post-groups in this study [2–4]. Robles and colleagues [1] mention effect size but do not report it.

We also would like to take a step back and challenge the field to think about pre- and post-assessment questions serving several purposes:

(1) To promote active learning
(2) To assess impact of the educational intervention based upon the learning objectives
(3) To provide insight into where learners are in their stages of behaviour change
(4) To further understand what in the education works for whom and why

We unpack each of these below.

Purpose 1: To promote active learning. Questions or quizzes can facilitate self-assessment and awareness of what one knows and where one needs more education. They are used in continuing education as a method for creating active learning – or participation from the learner in the learning process [5].

Purpose 2: To assess the impact of the educational intervention based upon learning objectives. Responses to questions may be used to assess the impact of an intervention. Different sampling frames may be used to assess the impact, which the authors mention based upon the analysis methods tested – missing data samples and complete data samples. Other methods for sampling and assessing impact include: randomised control trials and then variations on paired or cross-sectional, matched or non-matched comparison groups that are time-aligned or non-time-aligned. Research and evaluation methodologists have differing views on each of these, with a likely order from the most rigorous to the least rigorous being:

(1) Paired (pre/post), randomised sample, time-aligned data
(2) Paired (pre/post), matched sample, time-aligned data
(3) Paired (pre/post)
(4) Cross-sectional (post only), matched sample, time-aligned
(5) Missing data (pre/post), pooled
(6) Paired (pre/post), unmatched sample, time-aligned; paired (pre/posts) not matched sample, not time-aligned
(7) Cross-sectional (post only), unmatched sample, not time-aligned

What is common across all of these sampling frames and study designs is that statistical significance and effect size can be examined. Statistical significance, indicated by a *P* value, suggests if we can reject the null hypothesis of a statistical test's logic (for the

CONTACT Katie Stringer Lucero klucero@webmd.net Medscape, LLC, New York, NY, USA

© 2023 Medscape, LLC. Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

purposes of this paper is the variance observed is due to chance rather than the CME) [2]. For example, we can reject the null hypothesis that the difference in question scores pre vs. post, test vs. control, or pre vs. post in test vs. control, and so on is due to chance and not the intervention. Note, "control" in this context can mean randomised, matched, or unmatched.

Three issues that may warrant further discussion exist.

(A) Data which are not independent violate assumptions of the statistical tests used for pooled data analysis as depicted in this study.[1]
(B) Focus is on statistical significance is misleading to evaluators. If we think about the purpose of statistical significance values as giving us confidence that are results are due to the grouping variable (i.e. CME vs. no CME) which are associated with sample size, and the pooled analysis has larger sample size, then what value does that provide? Of course, the pooled analysis will be more likely to show a statistically significant result becaue the sample size is larger.
Focus on statistical significance and equating it with improvements in knowledge, competence, and performance does not align with the purpose of statistical significance. Effect size is mentioned, but it is not reported. Authors state: "the improvement in correct responses at the paired and pooled pre vs follow-up analysis, suggesting sample size did not influence our findings".[p. 3] It would be helpful to see a comparison of effect size of pooled vs. paired in % of correct responses.[1]
(C) C. We ask – what is the purpose of reporting percentage of correct responses pre vs. post education? It is multifaceted as mentioned above. Statistical significance is only one element to be examined using those data.

Purpose 3: To provide insight into where learners are in their stages of behaviour change. Models such as the Transtheoretical Model [6,7] and Awareness to Adherence Model [8] specify stages of behaviour change. We suspect different measured outcome profiles depending on the stage of change in which a learner is. The goal of CME as a field is to promote best-in-class patient care. The value of the data referenced in the article is not being fully credited. Pooled data do not allow one to understand learner level outcomes; evaluators and programme developers cannot tie pooled data back to where learners (clinicians) are in the behaviour change process. In order to leverage, the pre- and post-assessment data to understand stage of behaviour change for groups of learners, data need to be triangulated, and examined longitudinally, hence, paired. There needs to be an understanding of whether the learner had primarily a reinforcing or learning something new experience; that needs to be overlaid with self-efficacy, intent to change, and current practice to get a full picture of understanding the process of behaviour change.

Purpose 4: To further understand what in the education works for whom and why. It has been increasingly recognised that in social science research, programme evaluation studies that focus only on outcomes are not enough. Programme evaluation is the systematic collection and analysis of information related to the design, implementation, and outcomes of a programme, like the series of CE activities in the Robles study [1], for the purpose of monitoring and improving the quality and effectiveness of the programme. Programme evaluation is about understanding the programme through a routine, systematic, deliberate gathering of information to uncover and/or identify what contributes to the success of the programme and what actions are necessary to improve the programme [9,10]. Recently, evaluators and researchers have recognised reports on outcomes alone cannot address the complexities of the health professions context and have suggested alternative approaches [11–14]. In a recent article, Allen and colleagues have described issues with outcomes only studies and have suggested several approaches that could be used to go "beyond did it work" [15].

Robles et al. [1] allude to the potential limitations of pooled analysis for heterogeneous groups. In CME, groups can be quite heterogeneous depending upon which variables are considered. There may be variation in motivation to learn, volume of patients with whom the clinician sees for whom the content may be relevant, the speciality, profession, role in patient care, geographic location, baseline understanding of, skills related to, and confidence in applying the topic, and other factors that may limit the way one learns from the specific intervention. We cannot assume that because learners are of the same profession and speciality that they are a homogenous group. Assuming there is heterogeneity in most learners, another way to leverage pre- and post-assessment data is to be analysed with other variables to understand moderators and mediators from descriptive segmentation and multivariate analyses. For example, one may segment pre- and post-scores by pre-CME self-efficacy in ability to use the skills taught in an intervention. Those who are admittedly less self-efficacious may be more willing

to learn the content because they are admitting they may need education. If the pre- and post-scores and change from pre-to-post are similar for each group, then we know that self-efficacy going into the intervention is not a moderator.

Our concern in writing this letter is not simply to point out the shortcomings of the Robles study [1]. Rather, our purpose is to shine a light on an important issue (evaluation beyond outcomes) that needs to be addressed so the field of CME/CPD can move forward with quality studies that will contribute to improved clinician performance and patient health. Practically speaking, the study which prompted this letter may mislead programme developers, evaluators, and funders to rely on statistical significance as the primary indicator of success of an educational activity. It may also lead them to not consider the full value of the pre- and post-questions and their underlying data.

Disclosure statement

This study has been solely submitted as original research to the Journal of CME and is neither published nor is under consideration elsewhere.

Funding

There is no source of funding for this study.

ORCID

Katie Stringer Lucero http://orcid.org/0000-0002-3278-7878

References

[1] Robles JH, Harb KJ, Nisly SA. Paired or pooled analysis in continuing medical education, which one is better?. J Cme. 2023;12(1). doi: 10.1080/28338073.2023.2217371

[2] Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ. 2012 Sep;4(3):279–282. doi: 10.4300/JGME-D-12-00156.1

[3] Maher JM, Markey JC, Ebert-May D. The other half of the story: effect size analysis in quantitative research. CBE Life Sci Educ. 2013;12(3):345–351. doi: 10.1187/cbe.13-04-0082

[4] Kühberger A, Fritz A, Lermer E, et al. The significance fallacy in inferential statistics. BMC Res Notes. Mar 17 2015;8(1):84. doi: 10.1186/s13104-015-1020-4

[5] Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478–485. doi: 10.3109/0142159X.2011.565828

[6] Prochaska JO, DiClemente CC, Norcross JC. In search of how people change: Applications to addictive behaviors. Amer Psychol. 1992;47(9):1102–1114. doi: 10.1037/0003-066X.47.9.1102

[7] Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: Toward an integrative model of change. J Consult Clin Psychol. 1983;51(3):390–395. doi: 10.1037/0022-006X.51.3.390

[8] Pathman DE, KONRAD TR, FREED GL. The awareness-to-adherence model of the steps to clinical guideline compliance: the case of pediatric vaccine recommendations. Med care. 1996;34(9):873–889. doi: 10.1097/00005650-199609000-00002

[9] Durning SJ, Hemmer P, Pangaro LN. The structure of program evaluation: an approach for evaluating a course, clerkship, or components of a residency or fellowship training program. Teach Learn Med. 2007;19(3):308–318. doi: 10.1080/10401330701366796

[10] Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. Med Teach. 2012;34(5):e288–99. doi: 10.3109/0142159X.2012.668637

[11] Haji F, Morin MP, Parker K. Rethinking programme evaluation in health professions education: beyond 'did it work?'. Med Educ. 2013 Apr;47(4):342–351. doi: 10.1111/medu.12091

[12] Olson CA, Bakken LL. Evaluations of educational interventions: getting them published and increasing their impact. J Contin Educ Health Prof. 2017; 37(4):281–284. doi: 10.1097/CEH.0000000000000181

[13] Olson CA, Shershneva MB, Brownstein MH. Peering inside the clock: Using success case method to determine how and why practice-based educational interventions succeed. Article J Contin Educ Health Prof. 2011;31(Supplement 1):S50–S59. doi: 10.1002/chp.20148

[14] Olson CA, Williams BW. Principles of effective research in continuing professional development in the health professions. In: Rayburn W, Turco M, and Davis D, editors Continuing professional development in medicine and health care - better education, better patient outcomes. The Netherlands: Wolters Kluwer; 2018. pp. 363–384.

[15] Allen LM, Hay M, Palermo C. Evaluation in health professions education-Is measuring outcomes enough? Med Educ. 2022 Jan;56(1):127–136. doi: 10.1111/medu.14654