

Incorrect Analyses of Cluster-Randomized Trials that Do Not Take Clustering and Nesting into Account Likely Lead to p -Values that Are Too Small

Lilian Golzarri-Arroyo, MS,¹ J. Michael Oakes, PhD,²
Andrew W. Brown, PhD,¹ and David B. Allison, PhD¹

Childhood obesity is a serious and challenging problem. Aceves-Martins et al.¹ took on a difficult and important task when designing and executing a cluster-randomized controlled trial (cRCT) that had as its primary aims to help participants “increase both their fruit and vegetable consumption and their physical activity (PA) while reducing their sedentary behavior.” Secondary aims were to decrease anthropometric obesity indicators. Probative studies—studies capable of meaningfully advancing our knowledge about the truth or falseness of a scientifically testable proposition—are sorely needed in this domain.² Yet, for studies to be probative, they must also be rigorously designed, executed, analyzed, and reported.³

In this light, Aceves-Martins et al. are to be commended for registering their trial,⁴ publishing a protocol article with an analytic plan⁵ and stating “If necessary, we will guarantee public access to the full protocol, participant database and statistical code.” They further stated in their protocol article, “Generalised linear mixed models are used to analyse differences between the intervention and control groups and changes in primary and secondary outcomes from baseline to the end of the intervention. For the rest of the efficacy variables, we will use Fisher’s exact test for the categorical variables and Student’s t -test for the continuous variables”.⁵ However, the Aceves-Martins et al. article refers to different analyses stating that “Generalized linear models (GLMs) were used to analyze differences from baseline to the end of the study in the primary outcomes of the intervention and control groups. Repeated-measures GLMs were used to analyze the trends in BMI between the baseline and end-of-study values. McNemar tests were performed to analyze the

changes in the primary outcomes in the intervention and control groups over time.”¹ Thus, a first concern is unexplained deviation from the prespecified analytic plan. A greater concern is that the authors do not specify whether (and, if so, how) clustering and nesting were taken into account in any of these analyses.

The concern about clustering and nesting is a serious one. As the NIH tutorial on cRCTs [which they term group-randomized trials (GRTs)] states: “Positive ICC [intra-class correlation coefficient] reduces the variation among the members of the same group but increases the variation among the groups. As such, the variance of any group-level statistic will be larger in a GRT than in a randomized clinical trial (RCT). Complicating matters further, the degrees of freedom (df) available to estimate the ICC or the group-level component of variance will be based on the number of groups, and so are often limited. Any analysis that ignores the extra variation (or positive ICC) or the limited df will have a type 1 error rate that is inflated, often badly.”⁶

Given that there are only two clusters per treatment condition in the Aceves-Martins et al. article, the miscalculation of p -values can be severe if clustering and nesting⁷ are ignored. Aceves-Martins et al. do not report their df or ICC values (despite published guidelines on cRCT reporting specifying that ICCs should be reported for each outcome variable⁸). ICCs for the types of variables Aceves-Martins et al. studied in populations of children have been reported to range from ~ 0.013 to 0.091 by Gray et al.⁹ and from ~ 0.039 to 0.189 by Masood and Reidpath.¹⁰

To provide crude estimates of what the effect of properly accounting for clustering and nesting in the analyses, if

¹School of Public Health, Indiana University—Bloomington, Bloomington, IN.

²School of Public Health, University of Minnesota, Minneapolis, MN.

clustering and nesting have not been taken into account, we adapted a procedure from Hedges.¹¹ To do so, we calculated as though the sample size was constant across all four clusters, and that the p -values generated from the test statistics used with and without taking clustering and nesting into account would follow patterns similar to those generated from t -tests with and without taking clustering and nesting into account. We emphasize that these calculations offer only crude approximations. That said, under the conditions of this study, if the ICC was as small as 0.05, a two-tailed p -value of 0.01 calculated (mistakenly) without taking clustering and nesting into account would translate to a two-tailed p -value of ~ 0.29 if clustering and nesting were properly taken into account. Under these circumstances, even a two-tailed p -value of 0.001 calculated (mistakenly) without taking clustering and nesting into account would translate to a nonsignificant two-tailed p -value of ~ 0.18 if clustering and nesting were properly taken into account. Moreover, even this correction from Hedges may be too liberal as noted by others.¹²

Given the mentioned, we ask that the authors fulfill their pledge to “guarantee public access to the full protocol, participant database and statistical code” by publically sharing the deidentified raw data and code, and to clarify how their analyses were conducted. If they did not properly take clustering and nesting into account as appears to be the case, the authors should publish a correction with the data analyzed explicitly taking nesting and clustering into account and revising conclusions if necessary. We offer our assistance with updated analyses if the authors wish.

Acknowledgment

This study was supported, in part, by NIH Grant Nos. R25DK099080 and R25HL124208. The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization.

Author Disclosure Statement

D.B.A. reports grants from NIH, outside the submitted work, and D.B.A. has received personal payments or promises for same from American Society for Nutrition; American Statistical Association; Biofortis; Columbia University; Fish & Richardson, P.C.; Frontiers Publishing; Henry Stewart Talks; IKEA; Indiana University; Laura and John Arnold Foundation; Johns Hopkins University; Law Offices of Ronald Marron; MD Anderson Cancer Center; Medical College of Wisconsin; NIH; Sage Publishing; The Obesity Society; Tomasik, Kotin & Kasserman LLC; University of Alabama at Birmingham; University of Miami; Nestle; and WW (formerly Weight Watchers International, LLC). Donations to a foundation have been made on his behalf by the Northarvest Bean Growers Association. D.B.A. is an unpaid member of the International Life Sciences Institute North America Board of Trustees. D.B.A.’s institution, Indiana University, has received funds to support his research or educational activities from NIH; Alliance for Potato Research and Education; American Federation for

Aging Research; Dairy Management, Inc.; Herbalife; Laura and John Arnold Foundation; and Oxford University Press. D.B.A.’s prior institution, the University of Alabama at Birmingham, received gifts, contracts, and grants from the Coca-Cola Company, Pepsi, and Dr. Pepper/Seven Up. In the past 12 months, Dr. Brown has received travel expenses from University of Louisville; speaking fees from Kentuckiana Health Collaborative, and Rippe Lifestyle Institute, Inc.; and he has been involved in research for which his institution or colleagues have received grants from Dairy Management, Inc., NIH, and the Sloan Foundation. The other authors declare that they have no conflict of interest.

References

1. Aceves-Martins M, Llauredó E, Tarro L, et al. A school-based, peer-led, social marketing intervention to engage Spanish adolescents in a healthy lifestyle (“We Are Cool”—Som la Pera Study): A parallel-cluster randomized controlled study. *Child Obes* 2017;13:300–313.
2. Casazza K, Allison DB. Stagnation in the clinical, community and public health domain of obesity: The need for probative research. *Clin Obes* 2012;2:83–85.
3. Wood AC, Wren JD, Allison DB. The need for greater rigor in childhood nutrition and obesity research. *JAMA Pediatr* 2019;173:311–312.
4. ClinicalTrials.gov Identifier: NCT02157402. EYTO (European Youth Tackling Obesity): A Randomized Controlled Trial in Catalonia (Spain) (EYTO-SPAIN). [Clinical trial]. 2014. Available at <https://clinicaltrials.gov/ct2/show/NCT02157402> Last accessed June 4, 2019.
5. Llauredó E, Aceves-Martins M, Tarro L, et al. A youth-led social marketing intervention to encourage healthy lifestyles, the EYTO (European Youth Tackling Obesity) project: A cluster randomised controlled trial in Catalonia, Spain. *BMC Public Health* 2015;15:607.
6. National Institutes of Health. Research Methods Resources, Group- or Cluster-Randomized Trials (GRTs). [Research Methods Resources]. 2018. Available at <https://researchmethodsresources.nih.gov/grt.aspx> Last accessed June 4, 2019.
7. Oakes JM, Kaufman JS. *Methods in Social Epidemiology*. San Francisco, CA: John Wiley & Sons, 2017.
8. Campbell MK, Piaggio G, Elbourne DR, et al. Consort 2010 statement: Extension to cluster randomised trials. *BMJ* 2012;345:e5661.
9. Gray HL, Burgermaster M, Tipton E, et al. Intraclass correlation coefficients for obesity indicators and energy balance-related behaviors among New York city public elementary schools. *Health Educ Behav* 2016;43:172–181.
10. Masood M, Reidpath DD. Intraclass correlation and design effect in BMI, physical activity and diet: A cross-sectional study of 56 countries. *BMJ Open* 2016;6:e008173.
11. Hedges LV. Correcting a significance test for clustering. *J Educ Behav Stat* 2007;32:151–179.
12. VanHoudnos NM, Greenhouse JB. On the hedges correction for a t -test. *J Educ Behav Stat* 2016;41:392–419.

Address correspondence to

David B. Allison, PhD

School of Public Health—Bloomington

Indiana University

1025 E. 7th Street

Bloomington, IN 47405

E-mail: allison@iu.edu