Letter to Editor

Unsubstantiated conclusions from improper statistical design and analysis of a randomized controlled trial

Dear Sir,

The proper design and analysis of randomized controlled trials (RCTs) are essential for scientific progress identifying safe and efficacious interventions. We read the RCT recently reported by Jakhotia *et al*. with interest, which compared the effects of three forms of exercise and a control on health and fitness-related outcomes. Unfortunately, severe problems with the design and analysis of the study make the reported results highly questionable.

Specifically, the allocation of participants is not random. The randomization process generally consists of two steps: (1) Generating an unpredictable random sequence and (2) implementing the sequence in a way that conceals the treatment until subjects have been formally assigned to their groups. The protocol is described as assigning subjects "as they were recruited with n in group 1, n+1 in group 2,… and so on."[1] It is unclear if they put the blocks of individuals into groups or individuals themselves, but either method is not random if the order of enrollment is exclusively defining treatment allocation. With sequential enrollment, knowledge of the next participant assignment increases the risk of bias to include or exclude a potential participant based on demographic characteristics.[2] By violating both the principles of random allocation, the study reported cannot be properly referred to as an RCT.[2]

There are also concerns with how the statistical analysis was conducted. Table 2 of Jakhotia *et al*. reports baseline characteristics and tests for differences among the four groups using a repeated measures analysis of variance; this is an inappropriate test because there should only be data from one time point to consider. Furthermore, we attempted to replicate the tests using the provided summary statistics and calculated incompatible results [Table 1]. These errors, including an impossible $P > 1$ (perhaps a typographical error, as P values are probabilities and must be between 0 and 1), raise further concerns because these calculated baseline imbalances were the basis for the choice of test in later analyses. We found similar errors in Table 3 of Jakhotia *et al*., where the reported within-group mean difference and standard deviation, 95% confidence interval, test statistic, and P value were incommensurable with one another in several cases.

The objective of this study was to find differences in health-related outcomes between three exercise regimens and a control group, but the core of their conclusions focused on whether pairwise differences within the treatment groups were significantly different from baseline. This approach, where differences in nominal significance within-groups are interpreted as significant differences between the groups, is statistically invalid, producing false positives at a rate up to 87.5% for four groups.[3] That approach should never be used; instead we suggest performing *post hoc* exploratory pairwise comparisons with a multiple testing correction.

Table 1: Results from reported baseline "repeated measures" ANOVA and recalculated ANOVA *F*-tests for an overall difference between circuit, treadmill, Suryanamaskar, and control groups based on Table 2 of the original paper

Variable	Reported RMANOVA results		Recalculated ANOVA w/final group sizes (22/21/24/20)		Recalculated ANOVA w/initial group sizes (29/27/27/20)	
	F	P	$F_{3,83}$	P	$F_{3,99}$	P
Age	-	-	0.3040	0.8224	0.3464	0.7918
Height	-	-	0.9281	0.4310	1.1037	0.3514
Weight	-	0.0003	3.6757	0.0153	4.0894	0.0088
Body fat percentage	-	2.0073	4.9310	0.0033	6.2731	0.0006
Muscle mass	1.303	0.6175	0.5535	0.6472	0.6823	0.5649
Bone mass	0.684	0.5657	0.8666	0.4618	1.0736	0.3639
BMI	-	0.0007	3.6496	0.0159	3.8799	0.0114
Metabolic age	1.229	0.3082	2.4038	0.0733	2.8267	0.0425
Body water percentage	0.4344	0.7293	0.7908	0.5024	0.9513	0.4190
Visceral fat rating	2.685	0.0556	3.5623	0.0176	4.5715	0.0048
Estimated VO_2 max	-	0.0001	0.7395	0.5314	0.9550	0.4172
Upper limb endurance	-	0.0001	1.4566	0.2324	1.8285	0.1469
Lower limb endurance	-	0.0001	2.6988	0.0510	3.0633	0.0316
Sit and reach flexibility	-	0.0001	0.7068	0.5506	0.8421	0.4740

ANOVA = Analysis of variance, RMANOVA = Repeated measures analysis of variance, BMI = Body mass index

The title of this paper is also misleading; equivalency of treatments can only be properly determined through an equivalency trial,[4] which this study was not. Considering the compromised randomization, the irreproducible baseline statistics, the use of subsequent analytical strategies that depended on the faulty baseline statistics, and the inferences inappropriately based on within-group comparisons, we believe retraction of Jakhotia *et al.* is consistent with the recommendation of the International

Committee of Medical Journal Editors: "Errors serious enough to invalidate a paper's results and conclusions may require retraction".[5]

Acknowledgment

Supported in part by NIH grants R25HL124208, T32DK062710, R25DK099080, and P30DK056336. The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization.

Financial support and sponsorship

Supported in part by NIH grants R25HL124208, T32DK062710, R25DK099080, and P30DK056336.

Conflicts of interest

There are no conflicts of interest.

Brandon J George[1], TaShauna U Goldsby[1,2], Andrew W Brown[1,2], Peng Li[1], David B Allison[1,2,3]

[1]Office of Energetics, School of Public Health, University of Alabama at Birmingham, [2]Nutrition Obesity Research Center, University of Alabama at Birmingham, [3]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama, USA

Address for correspondence:
Dr. Brandon J George,
1700 University Blvd., Room 408, Birmingham, Alabama 35233, USA.
E-mail: brgeorge@uab.edu

REFERENCES

1. Jakhotia KA, Shimpi AP, Rairikar SA, Mhendale P, Hatekar R, Shyam A, *et al.* Suryanamaskar: An equivalent approach towards management of physical fitness in obese females. Int J Yoga 2015;8:27-36.
2. Dettori J. The random allocation process: Two things you need to know. Evid Based Spine Care J 2010;1:7-9.
3. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. Trials 2011;12:264.
4. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. JAMA 2006;295:1152-60.
5. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals [Internet]. International Committee of Medical Journal Editors. c2014. Available from: http://www.icmje.org/icmje-recommendations.pdf. [Last cited on 2015 Sep 08].

Access this article online	
Website: www.ijoy.org.in	**Quick Response Code**
DOI: 10.4103/0973-6131.171726	

How to cite this article: George BJ, Goldsby TU, Brown AW, Li P, Allison DB. Unsubstantiated conclusions from improper statistical design and analysis of a randomized controlled trial. Int J Yoga 2016;9:87-8.