Letter to Editor

Considering the design effect in cluster sampling

Yousef Alimohamadi[1,2], Mojtaba Sepandi[3,4*]

[1]Pars Advanced and Minimally Invasive Medical Manners Research Center, Pars Hospital, Iran University of Medical Sciences, Tehran, Iran
[2]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[3]Health Research Center, Lifestyle Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
[4]Department of Epidemiology and Biostatistics, Faculty of Health, Baqiyatallah University of Medical Sciences, Tehran, Iran

Dear Editor,

We read the valuable manuscript with the title: *Association of modified Nordic diet with cardiovascular risk factors among type 2 diabetes patients: a cross-sectional study* that published in *J Cardiovasc Thorac Res*.[1] In this manuscript authors have said "cluster random sampling methods were used to select participants. Throughout the five sectors of Isfahan, we randomly chose two centers (clinics) from each area. Based on previously calculated mean and standard deviations for BMI in this population, our target number of participants was 143".

We have some points about sampling method and sample size determination in mentioned manuscript. When cluster sampling is used the effect of intra-cluster correlation (ICC, or the strength of correlation within clusters) must be regarded for sample size calculation.[2] This effect called the design effect (Deff). It is a correction factor that is used to adjust the required sample size for cluster sampling. The required sample size, assuming a simple random sample (SRS), should be calculated, and then multiplied by the Deff.[3, 4]

Deff represents the degree of variance inflation that attributes to cluster sampling.[5] The value of Deff for prevalence estimation in a cross-sectional study depends on the average number of subjects sampled per cluster (n) as well as ICC[3]:

Deff$= 1 + p (n - 1)$

p=ICC

Suppose in a cross-sectional study, SRS is used to estimate the mean of a parameter. The variance of the sample mean is calculated as follows:

$$V_{SRS}(\bar{Y}) = \sigma^2/n$$

If instead of SRS, a cluster design is used with k clusters and m subject in each cluster, the variance is calculated as follows[5]:

$$V_{Cluster}(\bar{Y}) = (\sigma^2/k*m) * [1 + p(m - 1)]$$

$k \times m = $n

The expression $[1 + p (m - 1)]$ in the above formula is called the variance inflation factor (VIF). So in cross-sectional studies when the unit of sampling is a cluster, but the sample size is determined according to the SRS assumption, the determined sample size should be multiplied by Deff or VIF[3]:

Deff$=$ Variance$_{Cluster}$ /variance$_{SRS}$

n = n$_{SRS}$* Deff

Ethical issues
Not applicable.

Competing interests
Authors declare no conflict of interest in this study.

References
1. Daneshzad SE, Darooghegi Mofrad M, Saraf-Bank S, Surkan PJ, Azadbakht L. Association of modified Nordic diet with cardiovascular risk factors among type 2 diabetes patients: a cross-sectional study. **J Cardiovasc Thorac Res** 2018;10(3):153-61. doi: 10.15171/jcvtr.2018.25.
2. Rowe AK, Lama M, Onikpo F, Deming MS. Design effects and intraclass correlation coefficients from a health facility cluster survey in Benin. **Int J Qual Health Care.** 2002;14(6):521-3. doi: 10.1093/intqhc/14.6.521.
3. Carlin JB, Hocking J. Design of cross-sectional surveys using cluster sampling: an overview with Australian case studies. **Aust N Z J Public Health**. 1999;23(5):546-51. doi: 10.1111/j.1467-842X.1999.tb01317.x.
4. Otte M, Gumm I. Intra-cluster correlation coefficients of 20 infections calculated from the results of cluster-sample surveys. **Prev Vet Med**. 1997;31(1-2):147-50. doi: 10.1016/S0167-5877(96)01108-7.
5. Chow SC, Liu JP. **Design and Analysis of Clinical Trials: Concepts and Methodologies**. New York: John Wiley & Sons; 2008.

Please cite this article as: Alimohamadi Y, Sepandi M. Considering the design effect in cluster sampling. **J Cardiovasc Thorac Res** 2019;11(1):78. doi: 10.15171/jcvtr.2019.14.

*Corresponding Author: Mojtaba Sepandi, Email: msepandi@gmail.com