Review of: "Statistical Overview of Prevalence of Anaemia with Associated Socioeconomic and Demographic Factors in Nigeria"

Dawit Tesfaye[1]

1 Hawassa University

Potential competing interests: No potential competing interests to declare.

This manuscript tries to address the prevalence of anemia and its association with sociodemographic and economic factors by analyzing secondary data from the 2021 Nigerian Demographic and Health Survey. My comments focus on the method section and follow here.

-Most demographic and health survey studies use a complex sampling design that involves stratification, clustering, and weighting. Such studies violate the assumption of "independent observation," and standard logistic regression analysis by ignoring violation of this assumption results in false positive findings. Therefore, sufficient description should be provided about the sampling method of the original survey (the 2021 DHS) to ensure the statistical validity of the results. If it involves a complex survey design, please consider re-analyzing the data by complex sample analysis.

-Please describe the sample size, the N (%) of missing data in the dataset with possible reasons, and how you addressed missing data in the analysis (i.e., complete case analysis or other missing data handling methods).

-The dependent variable, described in the method section, lines 9 and 10, "…The prevalence of anaemia with associated socioeconomic and demographic factors in Nigeria," is not the correct way of writing the dependent variable. I suggest you describe the dependent variable clearly (i.e., anemia…), then describe how anemia was measured in the original study, and then the cut-off points that you have used to categorize anemia. Please revise.

-In the last sentences of the research and method section, I have seen three categories of anemia: normal, mild, and moderate/severe. However, the logistic regression analysis shown in Table 3.3 is not the appropriate choice to predict the outcome variable with three categories. Inline, despite multinomial regression being discussed in the method section, the results section presents nothing from the multinomial logistic regression analysis. Also, it is not clear why the authors mention the multinomial logistic regression in the method section? Please revise or clarify.

-Please describe how you checked the multicollinearity and interaction in your actual data, with findings

-I do not think "Statistical Overview" is relevant in the title; it does not reflect the content of the manuscript.

-Please follow the journal's guidelines to draft the manuscript.