

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

Michèle Dramaix-Wilmet¹

¹ Université Libre de Bruxelles

Potential competing interests: No potential competing interests to declare.

Title: What do the authors mean by "statistical overview"? The title is not clear and should be more in line with the analyses and results.

Introduction:

- The second sentence is not clear.
- There are statements without references; for example, the last two sentences of the 1st paragraph. In the first, it would be appropriate to cite some prevalences with references from all over the world, from low- and middle-income countries, and in the second, some references from studies on the consequences of anaemia. This is just one example; there are several others in the text.
- The references are not numbered, but some are cited with the author's first name and not their last name.
- What's 95% UI?
- End of the second paragraph. "59 months ²": What does the superscript "2" refer to?
- The introduction presents the situation in Nigeria and the factors associated with anaemia, but only one very old (12y) reference is cited.
- The reference of Kebede et al concerns anaemia in the world and sub-Saharan Africa; it should therefore be placed above, at the beginning of the introduction, to respect a usual pyramid structure: problem in the world, in Africa, in Nigeria, for example.
- The objectives are not clear at all. Some sentences of the last paragraph of the introduction seem to indicate that other groups than children aged 6-59 months will be studied, but the article concerns this group of children. The title doesn't specify that the article concerns children.

Research and methods:

- Could you, in a few sentences, explain the survey sampling design?

- The variable studied is anaemia, not prevalence. Prevalence is an epidemiological measure of disease frequency and not a variable.
- The factors that influence the status of the children are cited, but the present study in Nigeria seems to only consider socio-economic factors. Why this choice? Other factors not available in the database? However, the choice of the factors is not clear at all, and there is confusion between dependent and independent variables. Some information on the chosen factors should also be provided: how and why they were chosen, how they were categorised...
- The sentence "The independent variable of prevalence of anaemia with associated socioeconomic and demographic factors in Nigeria is the anaemia status of the children aged 6-59 months." is totally incomprehensible.
- Haemoglobin level was divided into 3 categories. How were the analyses carried out: considering the 3 categories or after dichotomisation (anaemia Yes/No)? It should be clarified.
- Logistic regression and χ^2 are well-known statistical techniques; there is no need to explain the theory and give formulas. Adequate references such as Altman and Hosmer & Lemeshow are preferred.

However, it would be essential to explain how the goodness-of-fit of the models and the absence of collinearity between predictors was verified, to specify how the multivariable models were constructed, and how the odds ratios and their confidence intervals were derived.

- The first paragraph of section 3 is also methodological. All elements of the statistical methodology must be grouped into a single paragraph at the end of the methodology, and the statistical package used must be included.
- DHS surveys have complex sampling designs, and to be correct, the analyses must take into account the characteristics of the sampling design (clusters, strata...). The statistical methodology in the DHS survey report probably explains how to do this.

Results

- The variables described in Table 3.1 don't correspond exactly to those cited in the methodology (e.g., Malaria, mosquito nets, dietary intake, etc.)
- The comment on the wealth index is unnecessary because usually, it is made up of quintiles and the % in each category is around 20%.
- What's the meaning of the sentence "The overall pattern suggests that there are differences in the distribution of the variables"?
- The heading "hypothesis of the study" in section 3.2 is simply the definition of statistical testing hypotheses and not a study hypothesis. There is no point in citing the null and alternative hypotheses involved in a statistical test.
- In Table 3.2, the % are necessary for the interpretation of the results.

- The sentence “A great number of under-five children who tested positive for anaemia reside in rural areas as a place of residence ($p < 0.001$)” needs to be rephrased because overall, there is a higher % of children from rural areas (Table 3.1). What’s important is to compare anaemia in rural and urban areas in terms of %.
- The results only show the results of the multivariable analysis. To have a better understanding of the results, the crude ORs are necessary. They should appear in Table 3.2 or in Table 3.3 instead of the unnecessary coefficients.
- The sentences: “The characteristics being considered include gender of the child, mosquito net usage, place of residence, wealth index, mother’s education, child’s age, and geopolitical zone. The results consist of the estimated odds ratio (OR) for the logit model. The P-values were included to validate whether or not each predictor variable in the model is significant. The influence of each covariate in the model is determined by its odds ratio for the adjusted model.” These sentences are methodological and not useful.
- The statistical significance should not be considered first. The sample is very large, and almost all tests will be significant even if the associations are weak. This is, for example, the case with the association between gender and anaemia. It’s significant, but the OR is just above 1, and the lower limit of the confidence interval is practically 1. Some associations are highlighted but not significant even with such a large sample. For example, the association between anaemia and place of residence is highlighted, but it’s not significant, and even the upper limit of the confidence interval is just slightly above 1.
- The methodology mentions multinomial logistic regression, but where is it used?

There is no discussion.

References

- The references cited are too few.
- Some references are incomplete (WHO Global Anaemia estimates, Kedebe et al.)

As it stands, the article is not ready for publication. There are things to clarify, and the analyses should be revised to take into account the sampling design. A structured discussion with good references should be added, and the introduction should be improved, in particular by adding references.