Reviewer’s report

Title: Nutritional status and HIV in rural South African children

Version: 4 Date: 15 October 2010

Reviewer: Carl J Lombard

Reviewer’s report:

1. The results reported seem to be a sub-group from a main survey. No detailed description on the sampling process of the main survey is provided and the only strata described detail is males and females. What were the age strata used in male and females? How was the overall sample size of 4000 determined and what precision was assumed for estimating HIV prevalence in a set of restricted strata (<60 months).

2. How was the children sampled within a village – was village the primary sampling unit?

3. What was the realisation of the sample in the different strata and how was the non-realisation handled? What about sampling weights? What about the non-response with respect to HIV?

4. The survey respondents age 12-59 months included in the analyses for investigating the impact of HIV on nutritional status has to take the complex sampling strategy into account. This is completely absent in the description of the statistical analysis and it seems that this aspect has been ignored. The statistical program Stata used for the analysis is geared for this but requires the proper setup of the sampling information etc. Estimation of prevalences, ratios, linear regression and logistic regression can be done with the right setup.

5. Impact of HIV status on nutritional status at community level. The current analysis done is incorrect with respect to the statistical analysis done as well as the conceptual thinking. You have only studied a single community and to assess the impact multiple communities will have to be studied in which the HIV prevalence varies and in which the important community co-variates are available. This is not a simple as doing an un-paired t-test as was done in the manuscript. The methods section, results section and discussion sections of the paper has to be cleared from this analysis. You can discuss the way to get an answer to the question at community level but your are limited by what you can do. You could look at the nutritional differential between villages with respect to HIV prevalence but the sample size of the strata is quite small.

6. Why are there gaps in the results for the categorized HAZ, WAZ, and WHZ tables? Surely the univariate results can be given. The univariate results in Tables 3-5 is just the estimated mean difference between the HIV subgroups of Table 2.

7. Categorizing covariates such as birthweight, household head age, mother’s age is ok for simple presentation of the data but the analysis should try and use
they as continuous outcomes. This allows one to investigate non-linear associations with the outcome. The age of the child was used in this way so the question one has to ask if the linear trend used in the analysis is appropriate. A simple non-parametric smoother of say WHZ on age using lowess will provide the evidence of how to proceed with this covariate in the models for example.

The study is an important one in view of the community setting but all of the current results need to take account of the statistical design of the study – a complex survey. They can contact the Biostatistics Unit of the MRC in Pretoria for specialized help with the analysis which I think they might need.