Reviewer’s report

Title: Identifying Children with Excess Malaria Episodes after adjusting for Variation in Exposure: Identification from a Longitudinal Study Using Statistical Count Models

Version: 1 Date: 9 March 2015

Reviewer: Amanda Ross

Reviewer’s report:

Ndungu et al have used a statistical model to identify children in two cohorts in Kilifi, Kenya who have had higher numbers of clinical episodes than the 95% centile expected on the basis of their age, exposure index, and calendar time.

Major comments

The discussion proposes the role of host factors as the reason for the excess malaria. It would be helpful also to mention other potential reasons (behavioural, economic and so on) for balance. Residual exposure is mentioned. The assertion (p13) “can be attributed to host factors” seems unjustified without further investigation.

Since the number of episodes in Kilifi has declined over this period, why does calendar time increase the risk of an episode? This may be due to using several covariates with polynomial regression, or to the limitations of local prevalence as a measure of transmission: the same value for prevalence may mean different things if the transmission is stable or if it has recently declined. Declining clinical immunity may also lead to different risks of an episode for the same prevalence over time.

The expected percentage of children above the 95% centile is 5%. There are 212/2463 (8.6%) over this cut-off, a factor of 1.72. The model does not fit perfectly so to say that the model is “optimized” (p2 abstract) is not quite correct: it is the best-fitting of several tried.

The “test of the utility of the ZIBR model” (p12) is limited to the application of a cut-off without adjusting for exposure. This does not seem to test the utility of the model so much as be a comparison to a crude cut-off. A number of things are different (the model, the covariates, the follow-up period) and so it is difficult to judge what is being tested. Since there is no external validation, you can only say that the children selected would be different. A relevant test would be to investigate the predictive power of the model: if the model detects some children as having higher than expected numbers of malaria episodes in a given time period, do they continue to have the same pattern for another time period?

Clustering by child is taken into account using robust variance estimates. There are other potential levels of clustering (household, area (relevant to the
prevalence measure), and cohort) which are not incorporated.

A brief description of the locally weighted prevalence measure would be helpful.

Interventions do not seem to be included in the individual’s risk. It is not stated if they are included in the aggregate prevalence measure. Bed net use is mentioned in the abstract, but does not appear in the model (Table). Is it possible you could be selecting children for not using bed nets in an area with high ITN coverage?

The first two sentences of the conclusions seem not to be rooted in the study results.

Minor points

Table 2: “No differences between the number of admissions” should be “no evidence of a difference”. The numbers, 88 and 56, are different.

The first paragraph of the results reads as if it should be in the methods section.

Final sentence of the Introduction (p5). “and allowing for random variation” – this seems to be tacked on to the sentence. It is the model, rather than the identification of children, which incorporates random variation.

The illustration of the poor fit of the Poisson model (p9) is fine, but from the literature many would not have assumed a Poisson model anyway.

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare I have no competing interests