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: 18 April 2015

Reviewer: Gabriela Gomes

Reviewer's report:

The authors build on previous work to explain the highly skewed distribution of the number of malaria episodes experienced by individual children living in the same environment. This is an important challenge in the epidemiology of malaria and other infectious diseases with high relevance for guiding research on disease mechanisms and design of effective control programmes.

The authors have previously described how the number of clinical malaria episodes was distributed in a data set of 373 children, and used Poisson regression together with a cut-off of >2 episodes above prediction to define a group comprising 21% of the children as having excess malaria (EM) (Mwangi et al 2008). Here they work with a larger data set (2463 children) and introduce zero-inflated distributions using an "exposure index" (Olotu et al 2012) to control for malaria exposure, and select children above the 95th percentile of observed minus expected numbers of clinical cases as EM (a group comprising approximately 9% of the children).

Major Compulsory Revisions:

As a test for the utility of the regression model proposed here, the authors compare applying a simple cut-off of an absolute number of clinical episodes >8 (which would have led to a similar number of EM) and conclude that around 15% of excess malaria would have been misclassified. I wonder where this >8 comes from and whether the authors could present a more robust test for their method.

The authors proceed to compare the identified EM group with a group of children matched for similar exposure but who had only experienced an average number of clinical episodes (AM), and conclude that EM children had higher parasitaemia and were less likely to be of the sickle cell trait genotype. Returning to the issue of testing the model, I am curious as to what the 15% of children misclassified by the arbitrary cut-off would show for these markers.

Minor Essential Revisions:

Page 2: In the abstract EM is used without defining.

Page 8: EM and AM are used here and a definition does not appear before page 11.

Page 15: "that these children 'that' fail" seems to have a 'that' too many.
Discretionary Revisions:
Over a series of papers the authors have built a research strategy to characterize heterogeneity in host susceptibility to infection that may inspire similar studies, not only in malaria but also other infectious diseases. Moreover, they have accumulated a large dataset that could be of great value not only for malaria but also for methodological research more generally if made publically available.

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests.