Reviewer's report

**Title:** Do patients with recurrent reported episodes of campylobacteriosis differ from those with a single disease event?

**Version:** 2  **Date:** 14 December 2010

**Reviewer:** Edith Gabriel

**Reviewer's report:**

Major compulsory revisions

In logistic regression models used by the authors, the outcome is the report of an episode of campylobacteriosis and explanatory variables are: previous episode of campylobacteriosis and characteristics of patients (either gender or age group or urbanicity). In the following, these models will be referred to as “simple”.

When patient characteristics are tested one by one, it is easy to assign a p-value to a characteristic. But, collectively, the p-values must be corrected for multiple testing, by some version of a False Discovery Rate (FDR) correction. Furthermore proceeding characteristic by characteristic omits the impact of other characteristics. Thus, it is possible for some variables to be significant with “simple” regression, but not with multiple regression. Conversely, it is possible that nonsignificant patterns in “simple” regression become significant in multiple regression. That's why multiple regression must be preferred.

A more compelling concern is that the model used by the authors is not correct. One aim of regression analysis is to estimate the effect of the causal variables upon the variable that they influence. In their manuscript the authors want to “describe the risk of having a recurrent episode of campylobacteriosis in relation to patient characteristics.” In order to investigate the associations between covariate information and recurrence a multivariable logistic regression model must constructed. The outcome must be the report of more than one episode (yes/no) and the explanatory variables must be age, gender, urbanicity and their interaction if any. This model does answer their question. It is not clear why the authors consider a model in which the outcome is the report of an episode (yes, no), ie why they add controls as they are only interested in cases and more particularly those with recurrent episodes. Furthermore, I am not convinced that one might really considered recurrence as an explanatory variable of the episodes. The use of case-case multivariate analysis above suggested is for example illustrated in Mullner et al.(2010).

The manuscript is an interesting piece of work, however this part must be made properly as corresponding results are the key to the study.

Epidemiology and infection.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests