Reviewer's report

Title: Using sequence data to identify alternative routes and risk of infection: A case-study of Campylobacter in Scotland

Version: 1 Date: 7 October 2011

Reviewer: Nigel P French

Reviewer's report:

Major essential revisions.
As a general comment I don't think the authors have been critical enough of the case-case approach, or outlined it's limitations sufficiently. If you had similar risk factor information on unaffected controls, you may find that both ruminant and poultry cases were associated with adults in urban areas in winter, but a stronger association between these factors and poultry cases would produce the (seemingly opposite) result observed.

My feeling is that greater care needs to be taken in the wording of the results and conclusions to avoid giving the impression that they are estimating either risk or relative risk. The limitations of case-case studies are outlined to some extent in the first such study using this method on Campylobacter case data (Gillespie et al 2002, EID) and this paper should be referred to in the present study in my view. In the abstract of that paper they remind us that using this approach "Exposures that are a risk for infection for both comparison groups might not be identified or might be underestimated by case-case analysis. Similarly, the magnitude or direction of population risk cannot be assessed accurately."

Minor essential revisions
Line 75. Provide references for multiple hosts.
Line 82 onwards. It seems odd that a lower deprivation score indicates higher deprivation? and this seems different to line 159. That said, it also seems incorrect to refer to it as a 'determinant' of 'human infection'. Surely all you can say is that there is an association between this index and the incidence of notified, confirmed cases of campylobacteriosis.
Line 86. Refer to original Wilson paper rather than software.
Line 137 onwards. Please indicate whether all the strain assignment in your study was carried out and reported in reference 4. Were all the STs isolated from humans in your study included in the Sheppard study?
I don't follow the sentence starting on page 139.
Presumably the structure method could assign STs that were only found in humans and not animals based on their allelic profiles?
Line 175. How many postcodes, how many Health Boards?
Line 231. Suggest adding some references to wild bird campylobacter and public health risks. Would be interesting to know what wild birds were sampled.
Discretionary revisions.
I would like to see the logistic regression model written down showing precisely how it was constructed.

**Level of interest:** An article whose findings are important to those with closely related research interests

**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