Author's response to reviews

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

Authors:

Paul R Bessell (prbessell@gmail.com)
Ovidiu Rotariu (ovidiu.rotariu@abdn.ac.uk)
Giles T Innocent (giles@bioss.sari.ac.uk)
Alison Smith-Palmer (alison.smith-palmer@nhs.net)
Norval JC Strachan (n.strachan@abdn.ac.uk)
Ken J Forbes (k.forbes@abdn.ac.uk)
John M Cowden (John.Cowden@hps.scot.nhs.uk)
Stuart WJ Reid (swireid@rvc.ac.uk)
Louise Matthews (Louise.Matthews@glasgow.ac.uk)

Version: 2 Date: 19 February 2012

Author's response to reviews: see over
Response to reviewers

The original comments by the reviewers are in italics and our response in plain text. Extracts from the manuscript are in a different font and new or altered text is in red in both the manuscript and the response.

Reviewer 1

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."

We appreciate that the distinction between a case-case analysis such as this and a study that aims to identify population level risk factors. As such we would not wish for the results to be interpreted as population level risk factors. However, in the original results we took pains to ensure that the base of comparison was clear, by using language such as: 'Compared to ruminant attributed cases, poultry attributed cases are more common in winter'.

Within the manuscript the following has been added to clarify that a case-case approach has been used.

The conclusions in the abstract have been reworded:
Rather than estimate relative risks for acquiring infection, our analyses show that individuals acquire C. jejuni infection from different sources have different associated risk factors.

In the discussion, the following paragraph has been added
By using a case-case approach this study did not seek to estimate population level risk of exposure. Rather this study analysed the subgroup of the population that has already been infected, with the principal risk factor being social deprivation [10]. Case-case analysis is a means of comparing risk factors within this sub-group of the population that has acquired infection [22] and has been employed elsewhere for comparing risk factors for infection between sources of C. jejuni [23]. As such, social
deprivation remains the principal population level determinant of infection with C. jejuni but these analyses demonstrate that this does not vary between sources of infection.

In the conclusions (in accordance with suggestions by reviewer 3):
Our results have demonstrated that over and above the previously demonstrated risk factors for infection at the population level [10], there are different risk factors for infection depending upon the sources of exposure to infection.

Minor essential revisions

Line 75. Provide references for multiple hosts.
Suitable reference have been inserted.

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. This is an error and has been corrected to say:
Previous studies have identified an association between human Campylobacter infection in Scotland and lower social deprivation score (indicating lower social deprivation) and being a child living in an area of lower population density [9].

Line 86. Refer to original Wilson paper rather than software.
This has been amended

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?
This was not a helpful wording in the original manuscript as the 2,420 related to the number of STs identified by Sheppard et al, but not all were isolated from human cases. Of those found in humans, all had source attribution data. This section has been reworded:
Each of 441 STs isolated from the 3,451 human cases of C. jejuni (372 cases that were infected with C. coli were removed from the analysis) was assigned a probability that the ST originated from poultry, cattle, sheep, wild bird and environmental sources as described in Sheppard et al. [6].

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?
This has been clarified in the correction to the previous comment.

Line 175. How many postcodes, how many Health Boards?
These figures have been added

Line 231. Suggest adding some references to wild bird campylobacter and public health risks. Would be interesting to know what wild birds were sampled.
Wild bird faeces were sampled in the study by Sheppard et al, so it is difficult to attribute these samples to a particular avian species. This has been clarified in the introduction to specify that it was wild bird faeces that were sampled. The sentence in the discussion has been amended to say:

This suggests that whilst wild birds are a reservoir there is little mechanism for human exposure, although exposure to preschool children in playgrounds has been suggested elsewhere [18].

**Reviewer 2**

**Minor comments**

*Page 8, line 144.* You refer to Ogden et al for the rational of lumping sheep and cattle into one category. Wouldn't it be better to first run the analysis with sheep and cattle as separate categories, and then if no differences were seen, lump them? With these data and this methodology this is not possible as there were only 16 STs that were of cattle origin with probability >0.95 that accounted for 18 cases and 12 sheep STs accounting for 96 cases. The vast majority of these ruminant cases were shared between cattle and sheep reflecting their similar digestive systems and often shared grazing.

*Page 5, line 77 and elsewhere.* It is bacterial isolates that are typed by multilocus sequence typing (MLST) into sequence types (STs). Strains are something else. Furthermore, one usually doesn't talk about ST groups: a ST is the distinct allelic profile from 7 loci. Increased levels of complexity are single-locus variants, double-locus variants and clonal complexes. I would suggest to look over this paragraph again. Similarly, the first line of the discussion states that MLST is a technology, but it is at most a genetic typing method. The text on line 77 has been corrected to:

**Reviewer 2**

**Minor comments**

*Page 8, line 144.* You refer to Ogden et al for the rational of lumping sheep and cattle into one category. Wouldn't it be better to first run the analysis with sheep and cattle as separate categories, and then if no differences were seen, lump them? With these data and this methodology this is not possible as there were only 16 STs that were of cattle origin with probability >0.95 that accounted for 18 cases and 12 sheep STs accounting for 96 cases. The vast majority of these ruminant cases were shared between cattle and sheep reflecting their similar digestive systems and often shared grazing.

*Page 5, line 77 and elsewhere.* It is bacterial isolates that are typed by multilocus sequence typing (MLST) into sequence types (STs). Strains are something else. Furthermore, one usually doesn't talk about ST groups: a ST is the distinct allelic profile from 7 loci. Increased levels of complexity are single-locus variants, double-locus variants and clonal complexes. I would suggest to look over this paragraph again. Similarly, the first line of the discussion states that MLST is a technology, but it is at most a genetic typing method. The text on line 77 has been corrected to:
Campylobacter can be classified by their allelic profile using Multi-Locus-Sequence-Type (MLST) typing techniques [6], which places isolates into specific Sequence Type (ST) profiles. Throughout the manuscript all mentions of ‘strains’ have been changed to ‘types’.

Reviewer 3

Minor comments

p. 8 description of analysed risk factors could be more clearly described in a Table.
We have considered this and feel that the description of the putative risk factors works best where it is as bullet points in the text and having to summarise this to fit it in a table would not benefit the situation. As neither of the other reviewers have made this suggestion we feel that we should retain the current format.

unassigned cases versus unassigned STs are easily misunderstood. (eg. p. 10, l. 204, table 1 and p. 13, l. 270. Use other some term for unassigned cases throughout the text.

We are reluctant to change the nomenclature from unassigned as this does describe the source assignment of the ST under this methodology. However, we appreciate that this is unclear, so we have clarified the nomenclature in the methods section:

Three separate case-case logistic regression analyses were carried out for all combinations of source of infection assignments. As this is a case-case analysis the group used for the base of comparison in the logistic regression are referred to as ‘controls’ despite them being incidences of disease:
1. Individuals infected with a poultry assigned type (cases) versus individuals infected with a ruminant assigned type (controls).
2. Individuals infected with an unassigned type (cases) versus individuals infected with a ruminant assigned type (controls).
3. Individuals infected with a poultry assigned type (cases) versus individuals infected with an unassigned type (controls).

Relative to ruminat cases- instead relative to ruminant assigned cases...
This has been corrected.

l. 206 imperfect throughout the text is/was...
This has been corrected.

l. 214 MLST technique istead of MLST technology; instead of species sources use host sources.
This has been corrected.

l. 237 showed instead of show, ....bovines are more common source in children...
showed that ruminant assigned types were more common in children in rural areas in summertime

I. 265, campylobacteriosis...
This has been corrected.

Conclusions:
Not paper but Our results/data has demonstrated... that infections sources vary.
This has been corrected to say:
Our results have demonstrated that over and above the previously demonstrated risk factors for infection at the population level [10], there are different risk factors for infection depending upon the sources of exposure to infection.

I. 284... For common genotypes....
Change to ‘genetic types’

I. 287, delete this work
Amended.