Author's response to reviews

Title: Risk factors for secondary transmission of Shigella infection within households: implications for current prevention policy

Authors:

   Lian Boveé (lbovee@ggd.amsterdam.nl)
   Jane Whelan (jwhelan@ggd.amsterdam.nl)
   Gerard Sonder (gsonder@ggd.amsterdam.nl)
   Alje P van Dam (avdam@ggd.amsterdam.nl)
   Anneke van den Hoek (avdhoek@ggd.amsterdam.nl)

Version: 6 Date: 31 October 2012

Author's response to reviews: see over
Dear Editor,

My co-authors and I would like to thank each reviewer for their time and expertise in critiquing this paper. Please find below a detailed response on a point-by-point basis, to each comment. We hope that the reviewers will find that their comments and suggestions are adequately reflected in the changes we have made to the manuscript, and that you will continue to consider this paper for publication.

Looking forward to hearing from you at your convenience,

Yours Faithfully,

Jane Whelan (corresponding author)
**Reviewer 1: Dirk Werber**

**Reviewer's report:**
Major compulsory revisions:

1. **Aim of the study**
   
   The authors’ state that the aim of their study was “to evaluate the appropriateness of current disease prevention guidelines in the Netherlands by determining the characteristics of primary case patients and of “high risk” household contacts that are associated with secondary transmission” High risk households were those where the primary patient was younger than 16 years or one or more contacts in the family were younger than 16 years.” However, the manuscript does not present secondary transmission rates in high-risk households where the primary case-patient is 16y or older. Consequently, the reader misses an evaluation of the appropriateness of considering these households as being at high-risk. Strictly speaking, if one was to evaluate the appropriateness of this disease prevention guideline, one would need to compare secondary transmission rates in high-risk households with that of low-risk households. I understand that this was not possible, but this limitation should at least be made explicit, or, preferably, the aim of the study should be rephrased. Are there any data on “household outbreaks” that might be used as an indicator for the risk of secondary transmission in low-risk households?

The designation of households with a case over 16 years as “low risk” for secondary transmission was based on research conducted in the Netherlands in 1999 (Vermaak et al) and our paper builds on this research. Prior to 2001, all contacts were routinely screened, irrespective of the age of the case. Vermaak et al. evaluated the appropriateness of this policy. Young age, of both the patient and the contact, were independently shown to be associated with a higher secondary attack rate, and secondary transmission among contacts over 16 years occurred in only 3 cases out of 416 contacts screened over 8 years, and all three without clinical symptoms. On this basis, it was decided that only contacts in “high risk” households (with children <16 years) would be screened thereafter (this has been the policy since 2000). Our study supports this finding to some extent: in households with a case over 16 years where selective screening was conducted (in those considered at highest risk, i.e. if the contact was symptomatic or was a food handler or carer), no secondary transmission occurred. Unfortunately (..or fortunately..) no outbreaks of shigella have been reported in households in the Netherlands in recent years (one occurred in 1992 involving MSM). We agree that it would be preferable if all contacts could have been studied, but this has not been possible since 2001.
We have rephrased the aim of the study, and amended the text as follows:

Pg. 4 – Last paragraph, background:
“The aim of this study was to determine the degree of secondary transmission within “high risk” households and the characteristics of primary case patients and their contacts that are associated with secondary transmission, and thereby to evaluate the appropriateness of the exclusion policies in relation to young children”.

Pg. 10 – Limitations of the study
firstly, we were unable to examine the risk of asymptomatic secondary transmission in low risk households, however among those at highest risk in these household who were screened (i.e. those who were symptomatic, or were care-workers or food-handlers) no secondary transmissions occurred.

We also describe in the discussion (1st paragraph, pg. 9) that Vermaak et al found only 3 asymptomatic secondary transmissions over an 8 year period in contacts of cases over 16 years.

2. Statistics/Study design
a.) What was the unit of analysis, and how exactly was household clustering taken into account in the (univariable and multivariable) analysis? Please make the model in that respect more explicit.

Because we had a large number of clusters (n=102) and the largest cluster was <5 percent of the total sample of contacts (n=337), we were able to use ordinary univariable and multivariable binomial regression models with clustered robust standard errors based on the Huber-White sandwich estimator (references below). This is a similar approach to using a generalised estimating equation with an independent correlation matrix. The unit of analysis was the individual contact (the denominator degrees of freedom are based on the number of observations, not the number of clusters). I have added references to the paper as follows, and amended the report of the statistical analysis.


Specifically, the approach used in STATA is summarized here:
http://www.ats.ucla.edu/stat/stata/Library/cpsu.htm

b) Analysis of the primary cases’ age is conducted using age as a categorical variable with three classes. The discussion however, often compares <6y to >6y implying that the two younger age strata
being collapsed. This seems justified based on the similar IRRs. It would be coherent though to collapse the categories in multivariable analysis as well.

We initially ran the analysis with age collapsed into two strata as you suggest. However, as the exclusion policy particularly relates to children attending pre-school (aged 0-3), and those attending junior classes in primary school (aged 4-5), and as the attack rate overall was highest in children aged 4-5 years, we concluded that, although mathematically nothing was added by presenting the age-groups separately, it was of interest to the reader, particularly given our research question.

We have clarified this further in the methods (Pg. 6):

“Age was classified in three age-groups based on school attendance: those aged 0-3 years attending pre-school, those in junior classes in primary school aged 4-5 years, and those aged ≥6 years who are fully toilet trained and capable of washing their own hands.”

... and have amended the introduction and discussion to refer to preschool and primary school children accordingly.

c) I haven’t understood the rationale for conducting a multivariable analysis that controls for factors associated with the primary case and the household. Could you kindly elaborate?

Based again on the research of Vermaak et al, we hypothesised that both the characteristics of the case and of the contacts within a household would determine the risk of secondary transmission. Specifically, we suspected that young age of the case in the household would increase the risk of secondary transmission, irrespective of the age of the contact, or the number of contacts in the household for example. For this reason, we included both individual level variables related to the contact (age, gender etc.) and contextual level variables related to both the household (e.g. size) and the case within the household (age, gender etc.). In the final model, the unit of observation is the individual contact (n=337). We have added this explanation to the statistical analysis, and have added a footnote to table 2 to clarify this.

c) How was “diarrhea in a household contact” defined (particular in respect to the temporal relationship with diarrhea in the primary case)?

Diarrhoea in a contact was self-reported – we did not have any objective definition, and contacts were simply asked on what day they had developed diarrhoea relative to the case. In total, 337 of 368 contacts were tested. For clarification, I have added the following to the text:

Pg. 5, Methods
Contacts were asked to report any symptoms experienced (diarrhoea, fever), and on what day they began relative to the primary case. For comparative purposes and to be consistent with previous
research[8], secondary infection was defined as laboratory confirmed *Shigella* infection in a household contact developed >1 day after the primary case...."

4. Nature of contacts’ relationship

*It is a crucial finding that most secondary cases were siblings or mothers of primary cases. This should/could have implications for targeted screening and this result should be discussed (and maybe also investigated in more detail, eg, what is the age-distribution of siblings in secondary and non-secondary cases). Furthermore, this result is consistent with studies that looked at secondary transmission in STEC O157 cases (eg., Werber et al, CID, 2006), and this should be mentioned as it strengthens this result and seems to make it somewhat generalizable.*

Of the 25 secondary cases, 7 (28%) were siblings (median age 4, IQR 2-12) and 7 (28%) were the mother of the case. An additional 6 were “other family contacts”, who were also young children (26%, n=6, median age 5.5, IQR 4-14). Statistically, siblings or mums or no more or less likely to become a secondary case that other relatives (Pearson Chi2=2.4383, p=0.656), and we did not find any difference in age between siblings who were secondary versus non-secondary cases (mean age 8.1 & 10.6 respectively, p=0.5016). Overall, the attack rate was highest in contacts aged 4-5 years, though being symptomatic was the most important factor in the multivariable model. When we looked at predictors only (leaving diarrhoea out of the model), we found that the age of the case was most important, rather than any attributes of the contact. In this context, our primary recommendation is to target screening at any contact who is symptomatic and additionally, at all contacts of cases, where the case is under 6 years old. We have added these findings to the results (pg. 8) and refer to this now in the discussion (ref. 19).

**Minor compulsory revisions:**

1. Hospitalizing the primary case patient

*(Immediately) hospitalizing the primary case patient is a suggested measure for reducing secondary spread in illnesses caused by STEC O157. Here, hospitalizing the primary case was associated with a reduced risk for observing a secondary case. This should be discussed.*

In this study, 4/73 contacts (5.5%) became secondary cases in households where the case was hospitalised, compared to 21/264 contacts (8%) in households where the primary case was not hospitalised. Statistically, however, this difference was not actually significant – table 2, univariable results.

2 Differentiating different types of cases

*A clear definition of primary, co-primary, and secondary cases is given in the Methods section. However, in the text this distinction of cases is often missing rendering some parts of the manuscript difficult to understand. Please amend where appropriate.*
We have reorganised some of the text to make this clearer. Throughout, primary cases are referred to as such (and not as patients for example), and secondary cases similarly.

**Discretionary Revision:**

1. **Introduction**

The frequency, incidence and case-fatality of shigellosis in the Netherlands, for which the prevention guidelines are formulated, would be informative to the reader and should be mentioned. This information is more important than the situation of shigellosis in developing countries.

As you suggest, we have amended the introduction to take account of the situation in the Netherlands, adding local references.

Pg 3. **Background:**

“In developed countries, *S. sonnei* and *S. flexneri* account for the majority of cases, and in the Netherlands about 75% of infections are imported, most frequently in the summer months. Nationally, 300-600 cases of bacillary dysentery are reported each year, yielding an annual incidence of approximately 3.2/100,000 population”
Reviewer 2: Koen De Schrijver

Minor essential revisions

1. Numbering authors 3 and 4?
Now corrected.

2. Background
2.1. Use of the word developed and developing countries rather industrialised or less industrialised
Also corrected.

2.2 Reference and selection of guidelines as example before going in details for the Netherlands. I wonder why you have chosen these countries. It looks rational to choose examples of procedures who are quite strict (Australia) and countries who have a more liberal approach. Then you introduce the procedures from the Netherlands. It looks more logic when you do it that way.

We have reorganised these references as you suggest. These countries were chosen for illustrative purposes only as the guidelines are accessible online and are in English, given that BMC is an English language journal.

2.3 S. sonnei always in italics and always with minuscule also for shigellosis
Corrected.

2.4 faecal or fecal
All spellings are now in British English.

2.5 IQR IRR abbreviations for?
All are now spelled out at first use.

3 References according the guidelines of BMC , .. , etc;
Also amended.

4. Figure and tables: Shigella or shigella, boydii, hospitalized and hospitalised, diarrhoea,
Reviewer 3: Isabel Oliver

The use of the term non-single in the definition of a high risk household is confusing. 'A single person' would usually mean someone not in a relationship in this case it presumably means someone who does not leave alone. I would suggest rephrasing to say 'a high risk household was defined as any household with more than one inhabitant including at least one child of <16 years.'

We agree that this could cause confusion, and have rephrased as you suggest:
Amendment Pg. 5: Household contact study
“A high risk household was defined as any household with more than one inhabitant including at least one child <16 years, where a primary case (of any age) stayed for at least one overnight, using shared toilet facilities, from the onset of symptoms to the date of notification to the Public Health Service.

The authors do not explain the rationale for selecting a date of onset of symptoms of >1 day after the date of onset of symptoms of the primary case, although they explain that using a period of >3 days would not change the results. It would be helpful to understand the rationale.

We wanted to be able to compare our findings to those of Vermaak et al, in their study conducted in 1999. To be consistent with this research, we used the onset of symptoms of >1 day after the date of onset of the primary case, though we suspected that some cases could in fact have been co-infections rather than true secondary transmissions. In the final outcome, using the more conservative 3 day duration, our findings did not change. The method section highlights our reason for selecting 1 day:

“For comparative purposes and to be consistent with previous research [8], secondary infection was defined as laboratory confirmed Shigella infection in a household contact developed >1 day after the primary case.”

The data are sound and the response is very good. The statistical methods need further clarification.

Reviewer 1 made similar comments in relation to the statistical methods. We have made a number of amendments to the statistical analysis (pg. 6) and have added appropriate references– hopefully this is now clearer. Please also refer to our response to reviewer 1 above.

There is no power calculation presented, presumably this is because the sample was identified through routine surveillance and it included all cases in the relevant period, however, the number of secondary infections is fairly low and it is possible that the study has insufficient power to detect some differences.
We didn’t make a power calculation because we were using all available data. It is true that there could be type 2 errors, but ultimately we did detect differences at both univariable and multivariable level. However, as you suggest, we now mention this as a limitation.

_In page 7 the authors describe that 13 cases were associated with one transmission and 6 cases with two. They do not describe if this was taken into account in the analysis or if they are any factors that may have been associated with more extensive transmission._

This was taken into account: in each of the 6 cases where 2 secondary transmissions occurred, the infections occurred within one household, and the clustered robust standard errors take account of the fact that it was the same case that infected the two contacts. We also initially performed a Poisson regression using the secondary infection count as the outcome, but we didn’t find any factor leading to more transmission to >1 contact. We do not discuss this in detail in the paper due to word count limitations, but have added if to the results (pg. 8):

“Thirteen primary cases were associated with one positive household contact and 6 primary cases were associated with 2 positive household contacts. No factor was identified that was associated with >1 secondary transmission”

_Was duration of symptoms and severity of symptoms in the primary cases considered as factors that may be associated with transmission? Shigella secondary attack rates are usually considered to be high._

Duration of symptoms for cases was not recorded, but days from date of onset of illness to date of notification was used as an approximate measure of duration of transmission risk (i.e. prior to receipt of hygiene advice from a health professional). This was not found to be significant (table 2). Whether or not the case was hospitalised was also used as a proxy for severity – ultimately this was not significant either at the univariable level: 4/73 contacts (5.5%) became secondary cases in households where the case was hospitalised, compared to 21/264 contacts (8%) in households where the primary case was not hospitalised (table 2).

_Is it possible that the ascertainment of positive contacts may have been reduced due to the delay in notification of cases and that cultures may have been negative by then? Although the losses to follow up are very small (6%) considering the need to submit a stool sample, it would be helpful to understand from the authors if they think that any bias may have arisen from differences between the contacts screened and those who were not. The conclusions seem reasonable although it would be useful to get the comments from the authors about the questions posed above and further discussion is needed regarding the limitations of the study._
It is certainly possible the positivity may have been reduced due to a delay in notification of cases. It is well recognised that isolation of *Shigella* in the laboratory is much more difficult than, for example *Salmonella*, and that therefore the bacterium is sometimes missed (a limitation not specifically of our study). According to our guidelines, the sample should be submitted within 24 hours. Unfortunately, we only had the date of onset and the date of notification – we didn’t know if the delay was because the patient didn’t present sooner, or if the GP had delayed in taking and submitting a sample.