Author's response to reviews

Title: Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium

Authors:

Niel Hens (niel.hens@uhasselt.be)
Nele Goeyvaerts (nele.goeyvaerts@uhasselt.be)
Marc Aerts (marc.aerts@uhasselt.be)
Ziv Shkedy (ziv.shkedy@uhasselt.be)
Pierre Van Damme (pierre.vandamme@ua.ac.be)
Philippe Beutels (philippe.beutels@ua.ac.be)

Version: 3 Date: 26 November 2008

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

Concerning: MS: 1140514843213838

Thank you for considering our manuscript entitled ‘Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium’ by Niel Hens, Nele Goeyvaerts, Marc Aerts, Ziv Shkedy, Pierre Van Damme, Philippe Beutels for revision for publication in BMC infectious diseases.

We are grateful for the comments raised by both referees and formulated a point-to-point reply to those comments raised by referee 1 and thank referee 2 for the assessment of the statistical techniques used. We have indicated the changes in the manuscript using the ‘track changes’ facility in word.

Your sincerely,

Niel Hens

Prof. dr. Niel Hens
Interuniversity Institute for Biostatistics and statistical Bioinformatics
Universiteit Hasselt
Agoralaan 1
B3590 Diepenbeek
Belgium, and
Katholieke Universiteit Leuven
Belgium
Tel: +32-11-268232
Fax: +32-11-268299
Web: www.uhasselt.be/niel.hens
Point-to-point reply

- We agree that we have written a rather long paper on a two-day population survey of contact behavior in Belgium. However, we believe this is rightfully so because of the scope of the paper which is to present the results of this survey in its full extent. We do realize that the structure of the paper was not entirely clear and therefore added an introduction to the methods section:

Since the contact survey contains a lot of information, we will, next to some descriptive statistics, use up-to-date statistical methodology to highlight different aspects from the data. We will first describe the use of association rules and classification trees to identify associations and types of contacts in the database. Then we will use weighted generalized estimating equations to properly model the number of contacts taking into account the correlation of contacts recorded by the same participant on two different days. Finally, we relate contact patterns to the spread of infectious disease using the next generation matrix. While this methodology provides more insight into how close contact infectious diseases are spread, we note that we do not aim to describe the stochastic nature of an emerging epidemic.

In this section we present some more detail on the methods used throughout this manuscript and refer to the literature for a more detailed description.

We refer to the POLYMOD research quite often because it puts this contact survey in perspective for several European countries. This paper presents a more in depth analysis of the Belgian survey which is different from the overall POLYMOD research in that we collected data over two days and included a holiday in the sampling period. We believe that it is not necessary to be familiar with the POLYMOD research and more specifically with the Mossong et al. (2008) paper, in order to understand the survey design for this particular survey. We believe our paper can be read as a standalone piece as well.

- We agree with the referee that different types of contacts are the main driving force behind the spread of different infectious diseases. We therefore added a few lines to the discussion:

We note that for different diseases, different levels of contact intimacy will play a role. For example, Mycobacterium tuberculosis will be transmitted during more intimate contact as compared to the common cold or influenza. Therefore, contrasting different types of contact data to serological/incidence data in the spirit of Wallinga et al. (2006) would substantiate empirical evidence about which type of contacts are most predictive for the spread of a specific infectious disease in a population.
Some recently submitted work (Goeyvaerts et al. 2008 and Ogunjimi et al. 2008), already referred to in this manuscript, use these data together with serological data on the varicella zoster virus in Belgium to illustrate that indeed some level of contact intimacy seems most predictive for the observed serological profiles. The methodology used is based on the work of Wallinga et al. (2006) and is believed to be beyond the scope of this paper. We now refer to this in the discussion too:

Moreover, both Ogunjimi et al. (2008) and Goeyvaerts et al. (2008) showed that contacts involving skin-to-skin touching and lasting at least 15 minutes were most predictive for the observed seroprevalence profile of varicella zoster virus in Belgium.

We indeed provide empirical evidence for what is commonly believed; that people who are intimate are more likely to touch each other and spend longer times together. We disagree that we would have been fishing only to show the obvious. Of course it is true that different contact intensities are relevant for different diseases, and the strength of the contact data we present here is that we not only qualitatively establish what the referee considers to be the obvious, we also quantify the contacts in terms of relative frequency and usual duration. Whatever the disease (as long as transmission is not predominantly sexual or more generally strongly heterogeneous in other aspects than age as was the case with SARS), the relevant contacts for their transmission were recorded in the diaries (including in households), and fitting procedures can be applied to select the contacts that most adequately explain the observed age-specific prevalence (and force of infection) for a particular infection. Therefore, it is not clear to us what exactly intimate contacts, propagating disease transmission, are and which generic hypotheses to test without a specific setting under consideration. Again we do not wish to zoom into any particular close-contact disease in this manuscript, but we do so in related work using the information we provide in the current paper. In order for other researchers to understand their meaning, strengths and weaknesses, thus enabling them to use these data for their own purposes, we believe the current paper is a useful addition to the scientific literature.

We agree that households, because of the intimate contacts, are ‘key’ to infectious disease transmission. Therefore, although the relationship to the contacted person was not registered in the diaries, we extracted the contacts with household members by comparing their age (as registered by each participant) with the age of the contacted persons as explained in a separate subsection in the data collection section:
Household Contacts

Although in order to keep the diary manageable, we intentionally did not record the relationship with the contacted persons, we extracted household-like contact data from the database by identifying those contacts with the same age as the registered ages of the household members (knowing that the contacts occurred at home and that an exact age was given, and not an estimated age range as for the ages of contacted persons from whom the exact age is unknown). We performed a sensitivity analysis with respect to the selected contacts of the same age.

We also added another figure to the paper (Figure 2) showing the distribution of different contact characteristics for contacts inside and outside the household. A sensitivity analysis, with respect to contacts with the same age, did not show any impact on the results. These findings illustrate that intimate contacts mainly take place inside a household, although there is a substantial portion of these contacts occurring outside the household. With these additional figures, we believe that an additional table with IQR for different types of contacts is redundant.

To answer the question where most contacts take place, we present the densities of the number of contacts by means of a boxplot in a new figure (Figure 3) for households, location, duration and frequency.

As a final descriptive statistic, \textbf{Error! Reference source not found.}, shows boxplots of the number of (close) contacts in and outside households (left upper panel); the number of contacts per location (right upper panel); the number of contacts for the different frequencies (left lower panel) and duration (right lower panel). Whereas most contacts inside a household involve skin-to-skin touching, this is not true outside the household. There is substantial heterogeneity in the number of contacts recorded at work/school and during leisure activities with a high number of contacts mostly observed at work/school. Although there are no apparent differences, the lower left panel in \textbf{Error! Reference source not found.} indicates that there is a larger number of contacts with longer duration. Note also that most contacts are frequent contacts whereas fewer contacts are yearly or first time contacts.

We would like to comment that all statistical techniques are briefly described in the paper and that it is not our goal to explain all methods in detail. We have provided several references for further reading. As pointed out in our first point-by-point reply, we added a small introduction to the methods section emphasizing the motivation of the methodology.
used. We believe that readers, not necessarily familiar with these methods, can interpret the results without further ado.

- We agree that this study has limitations because of random digit dialing and included in the discussion:

  Sampling in this survey was done by random digit dialing on fixed telephone lines and thus has its limitations with respect to the representativeness of the Belgian population, especially because of a substantial increase in the number mobile phone users without fixed telephone lines (Pickery and Carton, 2005).