Estimation of the serial interval of pertussis in Dutch households

Dennis E. te Beest\textsuperscript{a}, Donna Henderson\textsuperscript{b}, Nicoline A.T. van der Maas\textsuperscript{a}, Sabine C. de Greeff\textsuperscript{a}, Jacco Wallinga\textsuperscript{a}, Frits R. Mooi\textsuperscript{a}, Michiel van Boven\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a} Centre for Infectious Disease Control, National Institute for Public Health and the Environment, The Netherlands
\textsuperscript{b} University of Oxford, United Kingdom

**Abstract**

Increasing incidence has led to the re-appearance of pertussis as a public health problem in developed countries. Pertussis infection is usually mild in vaccinated children and adults, but it can be fatal in infants who are too young for effective vaccination (≤ 3 months). Tailoring of control strategies to prevent infection of the infant hinges on the availability of estimates of key epidemiological quantities. Here we estimate the serial interval of pertussis, i.e. the time between symptoms onset in a case and its infector, using data from a household-based study carried out in the Netherlands in 2007–2009. We use statistical methodology to tie infected persons to probable infector persons, and obtain statistically supported stratifications of the data by person-type (infant, mother, father, sibling). The analyses show that the mean serial interval is 20 days (95\%CI: 16–23 days) when the mother is the infector of the infant, and 28 days (95\%CI: 23–33 days) when the infector is the father or a sibling. These time frames offer opportunities for early mitigation of the consequences of infection of an infant once a case has been detected in a household. If preventive measures such as social distancing or antimicrobial treatment are taken promptly they could decrease the probability of infection of the infant.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

**Introduction**

Pertussis is a highly transmissible infectious disease caused by the bacteria *Bordetella pertussis* and, less frequently, *Bordetella parapertussis*. While pertussis infection is rarely severe in adults, it can be dangerous for infants who are too young for full vaccination (Guris et al., 1999; De Serres et al., 2000). Recent years have seen an increase in pertussis outbreaks in developed countries, with a simultaneous increase in the number of severe cases (van Boven et al., 2000; Grant and Reid, 2010; Cherry, 2012). It is customary for children to be vaccinated three or four times early in life. This has certainly contributed to the strong general decline in pertussis infection rates in developed countries, but at the same time it has become increasingly clear that vaccination does not protect against infection for life, and that infected vaccinated persons may act as a reservoir for transmission to infants (Wendelboe et al., 2005; de Greeff et al., 2010a).

An important question therefore is how best to protect infants that are too young to be vaccinated. To answer the question it is important to obtain insight into the transmission routes leading to infant infection, and the associated time scales of infection. Recent studies have uncovered the pivotal role of household members in transmission to the infant. In fact, siblings most commonly introduce the infection in the household, while mothers most often are the infector of the infant (Mooi and de Greeff, 2007; de Greeff et al., 2010b; Castagnini et al., 2012). These findings have led to pleas to add maternal vaccination, i.e. vaccination of pregnant women, to current vaccination programs (Mooi and de Greeff, 2007; Leuridan et al., 2011). An alternative possibility that has recently come to the fore is a cocooning vaccination strategy in which household members in families with a newborn are vaccinated (Kuehn, 2010). However, as vaccination is costly and does not necessarily allocate resources most cost effectively, it is of importance to examine alternative local measures such as contact reduction or the early administration of antimicrobial drugs in households with a suspected or confirmed infection.

In this study we estimate the (clinical onset) serial interval of pertussis, i.e. the time between symptoms onset of a case and its infector, using data from a prospective study on pertussis in households with an infant in the Netherlands (de Greeff et al., 2010b). The serial interval is determined by the incubation period of the infected person, i.e. the time between infection and symptoms onset of the infected person, the transmissibility of the infector person, and the relation between the latent and incubation periods of the infector.

\* Corresponding author. Tel.: +31 302744264.
E-mail address: michiel.van.boven@rivm.nl (M. van Boven).

http://dx.doi.org/10.1016/j.epidem.2014.02.001
1755-4365/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
person (Fine, 2003). Under mild conditions, the mean of the serial interval equals the mean of the generation time, i.e., the time between infection of a case and infection of its infectors (Svensson, 2007). Hence, the serial interval is closely tied to the speed with which an infection spreads between persons and in populations, and it is an important determinant of the controllability of an infectious agent (Fraser et al., 2004; Wallinga and Lipsitch, 2007).

Methods

Data

In 2006, the National Institute for Public Health and the Environment initiated a study of pertussis transmission within households. The families of infants aged less than 6 months and hospitalized with pertussis were asked to take part. Data was collected from all members of the participating household through laboratory procedures and a questionnaire. The laboratory procedures included a PCR and serological tests for pertussis on each participant.

Sensitivity and specificity of the serological test are 80% and 97%, using PCR- or culture-positive subjects as gold standard (de Greeff et al., 2010b). The questionnaire indicates age, relation to the infected infant, date of symptom onset (if any) for each household member, and vaccination status for all children younger than 13 years. In the Netherlands, infants are offered a primary vaccination series of 4 doses of whole cell DTP-IPV (since 1957), and an acellular pertussis preschool booster (since 2002). Vaccination coverage in the Netherlands has been high over the past decades (>95%; http://bit.ly/19Pcrl), and also in our study a small minority of persons either had unknown vaccination status or reported being unvaccinated (37/363, 10%). First day of symptoms was defined as first day of cough or first day of cough-preceding cold symptoms. A detailed description of the study is given in (de Greeff et al., 2010b).

Households in which cases were present that did not have a clearly defined first day of symptoms were excluded. This procedure removed 346 out of 560 households, leaving 114 households with a clearly defined primary case for analysis (de Greeff et al., 2010b; de Greeff et al., 2012). We further removed 24 uninformed households with a single case of pertussis, and 3 atypical households. Two of the atypical households had infected grandparents, and the third had twin infants. In the end, 87 households containing 241 infected persons (all with a clearly defined first day of symptoms) were included.

Analysis

Our data is broken into certain and uncertain serial intervals. We consider the difference in onset time between the first and second case in each household to be a certain serial interval. For later cases, we consider the differences between onset date of said case and all earlier household onset dates to be uncertain or possible serial intervals. For example, a household with three cases produces one certain serial interval (first to second case) and two uncertain or possible serial intervals (first to third case, second to third case).

In earlier analyses of influenza A outbreaks all serial intervals were assumed to arise from a common distribution. Here we use an extension of the algorithm which allows for differences between transmission routes (te Beest et al., 2013). We systematically investigate models which distinguish by person-type of infectors individuals and by person-type of infected individuals. Our notational conventions are such that, for instance, $M \rightarrow I$ represents mother-to-infant transmission, $S \rightarrow F$ denotes infection of the father by a sibling, and $A \rightarrow F$ denotes infection of the father by any other household person. Since our interest is mainly with the mean of the serial interval distribution, we assign a common variance parameter to all serial interval distributions, thereby reducing the number of parameters and avoiding overfitting the data (te Beest et al., 2013).

From histograms, the empirical serial interval distributions seem to be well-described by gamma distributions, so we choose gamma distributions to model serial intervals (generalized gamma distributions did not noticeably improve model fits; results not shown). We use a prior-based Expectation–Maximization algorithm to weigh the probabilities of the uncertain serial intervals. In our algorithm the prior probability that case $i$ has been infected by case $j$ is denoted by $\pi_{ij}$. Further, we let $m_i$ denote the number of possible infectors of case $i$. We assume that all possible serial intervals leading to infection of case $i$ have equal prior probability, i.e. $\pi_{ij} = (1/m_i)$ for possible infectors $j$ and $\pi_{ij} = 0$ otherwise. To give an example, both uncertain serial intervals for a third case in a household have prior probability 1/2, and all missing serial intervals of a fourth case have prior probability 1/3.

Our method of analysis follows Hens et al. (2012). Specifically, if we denote by $g(x|\theta)$ the probability of a serial interval of duration $x$ when the (discrete or discretized) serial interval distribution is specified by parameters $\theta$, then the probability that case $i$ has been infected by case $j$ is given by

$$p_{ij}(\theta) = \frac{g(x_{ij}|\theta)\pi_{ij}}{\sum_{k \neq j} g(x_{ik}|\theta)\pi_{ik}},$$

where $x_{ij}$ is the time between symptoms onset in case $j$ and case $i$. In case that a priori all potential infectors of a case have equal infection probability, i.e. $\pi_{ik} = \pi_{ij}$ for all possible infectors $k$ and $l$ of cases $i$ the above equation reduces to pure weighting with serial intervals (Wallinga and Teunis, 2004; Hens et al., 2012). This is the case for our analyses in the main text and Tables S1–S3. We keep the more general notation to stress how the analyses could be extended, e.g., by incorporation of alternative sources of information such as contact tracing information, spatial proximity information, or sequence data (Hens et al., 2012; Ypma et al., 2012, 2013; Teunis et al., 2013).

With the above preparations, the expected log-likelihood can be written as

$$E(\ell(\theta|\mathbf{x})) = \sum_{i=2}^{n} \sum_{j=1}^{n} p_{ij}(\theta) \log(g(x_{ij}|\theta)),$$

where it is understood that the primary case has label $i = 1$, that $p_{i1} = 0$ if case $j$ has onset of symptoms earlier than case $i$, and $p_{ij} = 1$ if case $j$ is the sole possible infecter of case $i$ (Hens et al., 2012).

The expected log-likelihood is maximized using an EM algorithm. An initial estimate $\theta^{(0)}$ of the parameters determining the serial interval distributions is used to calculate the expected transmission probabilities $p_{ij}(\theta^{(0)}, \mathbf{x})$ (19). Subsequently, the initial parameter estimates are updated by maximization of the expected log-likelihood in which the transmission probabilities are inserted. Formally, $\theta^{(1)}$ is calculated as

$$\theta^{(1)} = \arg\max_{\theta} \sum_{i=2}^{n} \sum_{j=1}^{n} p_{ij}(\theta^{(0)}, \mathbf{x}) \log(g(x_{ij}|\theta)) .$$

These steps are iterated until the parameter estimates converge. We repeat this process using various starting configurations to ensure that the parameters converge to values that maximize the expected log-likelihood.

The above formulation assumes no stratifications by person-type. However, it is easy to see how the above equations can be extended by letting the generation interval $g$ depend on the types $\tau(i)$ of individuals $i$, or the types of transmission pairs $(i,j)$ of individuals $i$ and $j$ (18). Specifically, the contribution to the likelihood of an observed difference $x_{ij}$ in onset of symptoms becomes $g_{\tau(ij)}(x_{ij})$. 
in the most general setting. Our main analyses stratify serial intervals either by person-type of the sender, in which case we assume $g \rightarrow g_{ij}(x_i)$, or by person-type of the receiver in which case we take $g \rightarrow g_{ij}(x_j)$. The person-types are infant (I), mother (M), father (F), or sibling (S), and so $\tau(i) \in \{I, M, F, S\}$.

Parameter confidence intervals are calculated using the chi-squared approximation of the profile likelihood (which is usually more accurate in terms of coverage probability than the well-known approximation based on Fisher information; McCullagh and Nelder, 1989), and model odds are based on AIC differences (Burnham and Anderson, 1998).

Results

An overview of the data is given in Fig. 1 and Tables 1, 2. Table 3 gives an overview of the analyses (full results are given in Tables S1–S2). Models that distinguish by type of the infecting person have a reasonable fit only if the stratification sets fathers apart from other infectors (Table S2). Models that stratify by the infected person perform reasonably well if a distinction is made between infants/siblings and other persons (Table S1). Within the main set of models, two models which allow for specific serial interval durations of sibling to infant and father to infant (M5 and M6) have the highest statistical support (Table 3).

Table 2

<table>
<thead>
<tr>
<th>Transmission route</th>
<th>Certain</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any→Any</td>
<td>87</td>
<td>239</td>
</tr>
<tr>
<td>Mother→Infant</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>Father→Infant</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Sibling→Infant</td>
<td>14</td>
<td>34</td>
</tr>
<tr>
<td>Infant→Mother</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Father→Mother</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Sibling→Mother</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Mother→Father</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Sibling→Father</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Infant→Father</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Infant→Sibling</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Mother→Sibling</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Father→Sibling</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Sibling→Sibling</td>
<td>2</td>
<td>20</td>
</tr>
</tbody>
</table>

Inspection of the models with high support shows that for most transmission routes the estimated mean serial interval is 19 days (model M5: 19.0 days, 95%CI: 15.3–21.1; model M6: 19.4 days, 95%CI: 16.0–21.1), and that transmission from sibling or father to the infant takes more than a week longer (model M5: 27.5 days, 95%CI: 22.7–32.9; model M6: 27.6 days, 95%CI: 22.7–32.9). Model M5, which provides a separate estimate of the serial interval of the mother-to-infant transmission route shows that the estimate is close to other transmission routes in the household (20.2 days, 95%CI: 15.5–23.1) but significantly shorter than sibling/father to infant transmission (model M3 versus model M5: $D = 7.0$, df = 1, $p < 0.01$).

Model M5 is attractive because it has high statistical support but also provides an estimate for transmission from mother to infant that is not confounded by other transmission routes. The fit of this model is investigated in detail in Fig. 2. Overall, the estimated gamma distributions give a good representation of the prior weighted serial interval distributions, and are in excellent agreement with the posterior serial interval distributions (Fig. 2). The most conspicuous difference between the prior and posterior weighted serial interval distributions is that a small number of uncertain serial intervals of very long durations (>100 days) have become highly unlikely by the analysis (Fig. 2). Quantile-quantile plots show no systematic deviations of the estimated distributions from the data (Fig. 3).

Discussion

Our analyses have provided quantitative support for the long-held belief that the serial interval of pertussis is long, in the order of several weeks (Anderson and May, 1992; Vynnycky and White, 2010). Our results have furthermore uncovered significant differences in the time scales of particular transmission events in the household. Specifically, while for most transmission routes the mean of the generation interval is approximately 19 days (95%CI: 15–21 days), an infection of the infant by the father or a sibling typically takes more than a week longer (28 days; 95%CI: 23–33 days). These differences are statistically significant and the means of the serial interval distributions can be estimated with fair precision. However, it should be noted that the estimated variances are large. As a consequence, variability in individual serial intervals can be substantial (Fig. 2).
Table 3
Attributes of models initially separated by type of the infected person. Symbols denote mother (M), father (F), and sibling (S). Shown are the estimated means and variances of the serial interval distributions, the maximized log-likelihood (\( \hat{\gamma} \)), the number of estimated parameters, and the model odds.

<table>
<thead>
<tr>
<th>Model</th>
<th>Serial interval (days) (95%CI)</th>
<th>Variance</th>
<th>( \hat{\gamma} )</th>
<th>Parameters</th>
<th>Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Any→Any: 20.9 (17.9,22.7)</td>
<td>213.1</td>
<td>−600.0</td>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>M2</td>
<td>Any→Infant: 23.1 (19.0,25.6)</td>
<td>209.0</td>
<td>−598.1</td>
<td>5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Any→Mother: 18.8 (14.5,22.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Any→Father: 20.2 (14.8,24.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Any→Sibling: 19.1 (13.4,23.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>Any→Infant: 23.1 (19.0,25.6)</td>
<td>209.3</td>
<td>−598.1</td>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Any→M/F/S: 19.3 (15.8,21.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>Mother→Infant: 20.2 (15.5,23.1)</td>
<td>199.7</td>
<td>−594.6</td>
<td>5</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Father→Infant: 26.7 (17.9,36.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sibling→Infant: 27.8 (22.1,33.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Any→M/F/S: 19.0 (15.3,21.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>Mother→Infant: 20.2 (15.5,23.1)</td>
<td>200.0</td>
<td>−594.6</td>
<td>4</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>S/F→Infant: 27.5 (22.7,32.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Any→M/F/S: 19.0 (15.3,21.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M6</td>
<td>S/F→Infant: 27.6 (22.7,32.9)</td>
<td>200.5</td>
<td>−594.7</td>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Other: 19.4 (16.0,21.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A number of key assumptions need scrutiny. First, we have based the analyses on a method that has been widely used to estimate serial intervals (Hens et al., 2012; te Beest et al., 2013; Vink et al., 2013). The method is intuitively appealing, but makes the simplifying assumption that the time to infection of a susceptible person does not depend on the number of infectious persons in the household in the at-risk period. In other words, there is no competition between infectious persons for infection of susceptible persons, and we assume that the phenomenon called generation interval contraction plays a minor role (Svensson, 2007; Kenah et al., 2008; Kenah, 2011). This was done in order to avoid introduction of estimands such as the latent period (i.e., the period from infection to a person becoming infectious) and the incubation period (i.e., the period from infection to onset of symptoms) that cannot be estimated directly from the data. As a consequence, our finding that infection of the infant by the father or a sibling takes longer than transmission through other pathways may be attributable to fathers and siblings being less infectious to their children than mothers (i.e., having less or less intense contacts, having lower bacterial loads, or both), to fathers/siblings having a more prolonged latent period than mothers, or to fathers/siblings becoming...

Fig. 2. Histograms of the serial interval distributions. Person-type stratifications are as in model M5 (Table 3). Black, prior-weighted serial interval distribution; red, posterior-weighted serial interval distribution; blue, fitted gamma distribution.

Fig. 3. Quantile-quantile plots comparing observed data to the fitted gamma distributions.
symptomatic earlier after infection than mothers. If direct evidence were available on the infectiousness over time of infected persons (e.g., by bacterial culturing or PCR of tracheal swabs), one could envisage meaningful extensions of the methods employed here to relate the onset of symptoms to the moment of infection and the onset of the infectious period [Kenah, 2011; Cauchemez and Ferguson, 2012].

Second, the analyses are based on the premise that serial interval durations are independent of the households in which the transmission events occur. In essence, this amounts to assuming that there is no household clustering of serial interval durations, i.e., some households have shorter serial interval durations than expected by chance and others having long durations. This was done for simplicity, and since our models seem to be able to adequately capture variation in the observations (Figs. 2 and 3). Moreover, tabulations of the data by household and cluster size (Fig. 1) did not reveal systematically deviant patterns. Nevertheless, it is conceivable that a model which explicitly includes the possibility of household factors to modulate serial interval durations would yield an even better fit to the data. We have explored whether there is an impact of cluster size on estimated serial intervals (as in te Beest et al., 2013), and found no differences between small and large clusters (results not shown). Another possibility that we have not explored here would be to include random effects at the household level. This would be an interesting avenue for further development of the methods, but we believe that for the current data the potential of such extensions to significantly impact the estimates is small.

Third, we have assumed that there is a single introduction in the household, and that all subsequent cases result from the infection chain arising in the household. Although pertussis transmission does not exclusively take place within households, it is clear that households are the most important setting for infection of infants in their first months of life [de Greef et al., 2010b; de Greef et al., 2012]. Furthermore, at the time of the study there was no evidence of sustained community transmission, and the patterns of infection are remarkably consistent across households, with only a minority of secondary cases arising in the first week after onset of symptoms in the primary case (Fig. 1). This suggests that the potential of multiple introductions in the household to impact the results is small. This is corroborated by a sensitivity analysis in which 12 observed short serial intervals of less than 7 days were removed, resulting in mean serial interval estimates that are close to the ones reported here (Table S3).

Earlier analyses have found that siblings most often introduce pertussis in the household, and that mothers are the most common source of infection of the infant [de Greef et al., 2012]. Moreover, pertussis infection in siblings and adults appears to be less frequently asymptomatic than hitherto believed [de Greef et al., 2010b]. If these findings are true in general it would open prospects for household interventions aimed at protecting the infant from infection, or at providing early mitigation of the consequences of infection. For instance, social distancing after a household member has been found infected could potentially decrease the probability of infection of the infant. More importantly, systematic prophylactic use of antimicrobial agents in infants with an infected or suspect household member could contribute to decreasing the probability of infection of the infant [Granstrom et al., 1987; De Serres et al., 1995; Halperin et al., 1999]. It could also serve to provide effective mitigation of the sequelae of an infection of the infant, before the paroxysmal stage of infection has set in [Altunajii et al., 2005]. Such a targeted intervention strategy would probably be more cost-effective than blanket vaccination of mothers or mothers-to-be, but it would require prompt recognition of a pertussis infection in the household by parents with a young infant.

Acknowledgments

We thank the Oxford University International Internship Programme for their role in arranging the visit of Donna Henderson to the National Institute of Public Health and the Environment (RIVM), and Niel Hens and Annemie Vink for helpful comments on a draft of the manuscript. This study was supported by RIVM project S/210096.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.epidemi.2014.02.001.

References


Estimating the generation interval of influenza A(H1N1) in a range of social settings. Epidemiology 24, 244–250.


