Epidemics 7 (2014) 1–6

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Estimation of the serial interval of pertussis in Dutch households

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

[a] Centre for Infectious Disease Control, National Institute for Public Health and the Environment, The Netherlands
[b] University of Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 24 September 2013
Received in revised form 10 January 2014
Accepted 6 February 2014
Available online 18 February 2014

Keywords:
Pertussis
Serial interval
Maternal vaccination
Household transmission
Expectation–Maximization

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.

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

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 infector (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/19JPcri), 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 uninformative 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 infector individuals and by person-type of infected individuals. Our notational conventions are such that, for instance, $M \to I$ represents mother-to-infant transmission, $S \to F$ denotes infection of the father by a sibling, and $A \to F$ 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 algortihm the prior probability that case i has been infected by case j is denoted by π_{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 $g(x|\boldsymbol{\theta})$ the probability of a serial interval of duration x when the (discrete or discretized) serial interval distribution is specified by parameters $\boldsymbol{\theta}$, then the probability that case i has been infected by case j is given by

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

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. $\pi_{ik} = \pi_{il}$ 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

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

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

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

$$\boldsymbol{\theta}^{(1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \quad \sum_{i=2}^{n}\sum_{j=1}^{n} p_{ij}(\boldsymbol{\theta}^{(0)}, \mathbf{x}) \quad \log(g(x_{ij}|\boldsymbol{\theta})) \quad .$$

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 $\tau(i, j)$ of individuals i and j (*18*). Specifically, the contribution to the likelihood of an observed difference x_{ij} in onset of symptoms becomes $g_{\tau(i,j)}(x_{ij})$