Mathematical Statistics
Stockholm University

Modeling the spread of penicillin-resistant *Streptococcus pneumoniae* in day-care and evaluation of intervention

Mikael Andersson
Karl Ekdahl
Sigvard Mölstad
Kristina Persson
Hans Bertil Hansson
Johan Giesecke

Research Report 2001:9

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.matematik.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2001:9**,
http://www.matematik.su.se/matstat

Modeling the spread of penicillin-resistant *Streptococcus pneumoniae* in day-care and evaluation of intervention

Mikael Andersson[1,2], Karl Ekdahl[2], Sigvard Mölstad[3], Kristina Persson[4], Hans Bertil Hansson[4] and Johan Giesecke[2]

August 2001

Abstract

In 1995, a disease control and intervention project was initiated in Malmöhus county in southern Sweden to limit the spread of penicillin-resistant pneumococci. Since most of the carriers of pneumococci are preschool children, and since most of the spread is believed to take place in day-care, a mathematical model, in the form of a stochastic process, for the spread in a day-care group was constructed. Aspects as seasonal variation and size of the day-care group were particularly considered. The model was the used for comparing results from computer simulations without and with intervention. Results indicate that intervention is highly effective in day-care groups with more than ten children during the second half of the year.
Keywords: Epidemic model; Stochastic process; Infectious disease; Pneumococci; Disease intervention

[1]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: mikaela@matematik.su.se. Financial support from Karolinska Institute is gratefully acknowledged.
[2]Swedish Institute for Infectious Disease Control, Karolinska Institute
[3]Unit of Research and Development in Primary Care, Mjölby
[4]Regional Centre of Communicable Disease Control, Malmö

1 Introduction

Streptococcus pneumoniae, also known as pneumococcus, is a very common bacterium all around the world, especially among young children. Mostly, it can be found in the upper respiratory tract, where it usually colonises its host without resulting in a clinical infection [1]. It is more common in developing countries, where as many as 85 % of young children can be colonised [2]. In developed countries, these figures are generally much lower. In Sweden, the prevalence among adults has been approximately 3 % [3], and among children in the range 15-50 % depending on age and frequency of contacts with other children [4].

Since most colonised individuals are preschool children and transmission of bacteria requires fairly close contacts between individuals, it seems that the environment where most of the spread takes place is day-care centres. Several studies show that children who attend day-care are more likely to be colonised than those who do not [5, 6, 7].

As the name suggests, pneumococci are the most common cause of pneumonia. More common illnesses in children are inflammation of the ear, acute otitis media (AOM) and upper respiratory tract infections (URTI). However, the risk of developing symptoms if one is colonised by pneumococci is actually rather low, somewhere around 15 % [8]. Nevertheless, since colonisation is so widespread, practically everyone will suffer from some of these illnesses during childhood, sometimes several times.

Up to about thirty years ago, all pneumococcal strains were highly susceptible to penicillin, which therefore was a very effective remedy against all infections caused by the bacteria. Since then, several strains have developed resistance against penicillin and other antibiotics, and have been spreading in most parts of the world [9]. In some European countries, particularly in Spain and Hungary, the proportion of resistant strains is as high as 40-70 % [10, 11].

In Sweden, this proportion was for a long time much lower in the whole country; approximately 3 % [12]. However, some studies in 1992 showed an increase in Southern Sweden to a level around 10 % [13, 14]. Therefore, a disease control and intervention project, the *South Swedish Pneumococcal Intervention Project* (SSPIP), was initiated in 1995 in Malmöhus county, the southernmost county in Sweden, with the aim to halt the spread of penicillin-resistant pneumococci with minimal inhibitory concentrations (MIC) of penicillin ≥ 0.5 mg/L (PRP) [15].

One of the most important measures within this project is to identify PRP colonised individuals as early as possible and, for children in day-care, take them out of the day-care group for the duration of carriage. In Sweden, it is recommended to obtain bacterial samples in cases of therapeutical failure of penicillin in the treatment of AOM or URTI, and culture them to detect possible colonisation with PRP. Family members, siblings and other close contacts of a patient with a positive culture of PRP, a so called index case, are sampled and all positive cases are isolated. In particular, if the index case attends day-care, all children and staff in his or her group are sampled (sometimes even the whole day-care centre) and those found to be PRP-positive are isolated at home until proven free of the bacteria.

During the first four years, neither a significant increase nor decrease in the proportion of PRP in Malmöhus county was detected [15]. The question is: Does this mean that the spread has actually stopped, i.e. would there have been a significant increase in absence of the intervention, or would we have seen the same trend anyway? One way of investigating this would be to find another comparable region with a similar proportion of PRP, where no intervention measures were taken, and see what happened there. Unfortunately, there is hardly any such region elsewhere in Sweden, since the prevalence of PRP is much lower outside Malmöhus county. Outside Sweden, demographic and other socioeconomic conditions are probably too different for reliable comparisons.

Another approach would be to construct a statistical model for the spread of pneumococci in a population and estimate the efficiency of the intervention measures [16, 17]. However, since dynamics of human populations and epidemiology of pneumococci are quite complex and not even fully understood, it seems to be very difficult to obtain a reasonably reliable model on the society level. Therefore, we will limit ourselves to a much smaller and more homogeneous population, which, as mentioned earlier, plays an important role in the spreading process, namely a day-care group. The large amount of data on pneumococcal carriage in day-care children available from the SSPIP provides good estimates for some of the parameters in the model.

The aim of the study is hence to use a model for the spread of pneumococci in the day-care setting to estimate the magnitude of the effect of the intervention introduced in the *South Swedish Pneumococcal Intervention Project*.

2 The model

The objective is to construct a model for the spread of pneumococci within a day-care group, consisting of a certain number of children. Let us therefore consider a fixed, closed population of n individuals. The assumption that the population is closed means that no individuals are entering nor leaving and that bacteria are only transmitted between individuals in the population. Since the duration of an outbreak is often comparatively short, the first assumption seems quite reasonable. However, the second assumption implies that a child is exposed to bacteria only by children in the same day-care group, which underestimates the overall exposure, e.g. to other day-care groups at the same day-care centre and to siblings. This underestimation is probably not substantial, since transmission of pneumococci bacteria requires quite close contacts, which are more frequent between children in the same day-care group than in other circumstances. Also, the primary aim is to study the dynamics *within* one single day-care group.

We assume that all individuals may be in any of the three states **S**usceptible, **C**arrier or **I**nfected. Being **S**usceptible means that an individual is completely uncolonised by bacteria, but may become colonised by having close contact with another colonised individual. If bacteria are transmitted, the individual becomes a **C**arrier, which means that he or she is colonised by bacteria, most commonly in the upper respiratory tract, without developing a clinical infection. After a short incubation period, a carrier may develop symptoms like

AOM or URTI, and thus become **I**nfected. Note that a carrier does not necessarily develop symptoms at all. It is actually more common to remain asymptomatic during the whole colonisation period.

Eventually, all carriers and infected individuals will recover and then become susceptible again. One might expect that a recovered individual would become immune, at least for a while, but the immunity to pneumococci in small children is very poor for most serotypes [8]. The individuals may acquire some partial immunity for a short time after recovery, but for simplicity we have not considered this.

Hence, there are four possible transitions between states:

1. From susceptible to carrier.
2. From carrier to infected.
3. From carrier to susceptible.
4. From infected to susceptible.

To complete the model, we need to specify the exact dynamics of these transitions. Since the population in question is rather small (n is typically in the range 5-25 children), we have chosen to describe the transitions as *stochastic processes*. In general, when considering dynamics of large populations, it is common to study purely deterministic models in terms of differential equations. This works well because all random phenomena on the individual micro-level disappear on the population macro-level. In our situation, individual variation probably plays an important role in the spreading process. To simplify things, we will only let transitions take place at discrete time points rather than continuously, one week seems to be a suitable time unit in this context. Then we do not have to consider weekends and we still manage to capture the rather slow colonisation cycle of pneumococci bacteria.

Let us first introduce some more notation. Let t denote the number of weeks since the bacteria was introduced in the population. Then we can define the three stochastic processes

$$\begin{aligned} S(t) &= \text{the number of susceptible individuals at time } t, \\ C(t) &= \text{the number of carriers at time } t, \\ I(t) &= \text{the number of infected individuals at time } t. \end{aligned}$$

We also need to keep track of the actual times when the carriers in the population were colonised. Therefore, we also introduce the process

$$C(t,s) = \text{the number of carriers at time } t \text{ that were colonised at time } s.$$

Since this is really a more detailed picture of the number of carriers, we get the relation

$$C(t) = \sum_{s=0}^{t} C(t,s).$$

Moreover, since preschool children often suffer from other infections and illnesses besides pneumococcal infection, some of them may be absent in any given week. We will describe the mechanism behind this in more detail later. At this point, we just introduce the new processes

$$\begin{aligned} S^*(t) &= \text{the number of susceptible individuals present at time } t, \\ C^*(t) &= \text{the number of carriers present at time } t, \end{aligned}$$

without making any further assumptions.

Finally, we also need four processes to keep track of the numbers of individuals that make each of the four transitions at each time interval.

$$\begin{aligned} X(t) &= \text{the number of susceptible individuals that become carriers at time } t, \\ Y(t) &= \text{the number of carriers that develop symptoms at time } t, \\ Z(t) &= \text{the number of carriers that recover at time } t, \\ U(t) &= \text{the number of infected individuals that recover at time } t. \end{aligned}$$

Again, we need a more detailed version of one of the processes above, namely

$$Z(t,s) = \text{the number of carriers recovering at time } t \text{ that were colonised at time } s.$$

Now, let us focus on the first and most important transition from susceptible to carrier. At time t there are $S^*(t)$ susceptible individuals and $C^*(t)$ carriers present in the day-care group. During the time interval between t and $t+1$, every susceptible individual is exposed to $C^*(t)$ carriers. Since a day-care group of children is a rather close and homogeneous group of individuals, we assume that the risk of transmission of bacteria from a carrier to a susceptible individual during one week is the same for all such pairs of individuals and denote it p. Furthermore, we also assume that all transmissions occur independently of each other. This means that in order to avoid being colonised, each susceptible individual has to avoid exactly $C^*(t)$ "attacks" and the probability for this can be expressed as $(1-p)^{C^*(t)}$.

This means that the number of susceptible individuals that will be colonised can be regarded as the result of $S^*(t)$ independent random trials, where the probability of "success" at each trial is $1-(1-p)^{C^*(t)}$. In other words, $X(t)$ is a binomially distributed random variable with parameters $S^*(t)$ and $1-(1-p)^{C^*(t)}$, or in mathematical notation

$$X(t) \sim \text{Bin}(S^*(t); 1-(1-p)^{C^*(t)}).$$

So far, we have assumed that the weekly transmission risk p is constant, which does not appear to be very realistic. One important aspect to consider is seasonal variation. Available data strongly suggest that the spread of pneumococci is much more frequent during autumn and winter than otherwise during the year. Exactly what this seasonal pattern looks like is hard to say without analysing available data more thoroughly. We will do that in the next section, but for now we just assume that there is some overall spreading intensity for each week of the year. We denote this λ_w, where w is the number

of the week from January 1st ($w = 1, 2, \ldots, 52$), and assume that p is proportional to this intensity.

Another aspect that may affect the transmission risk is the group size n. Having a fixed p for all group sizes implies that contacts between any two individuals are as close and frequent in a group of, say, five children as in a group of thirty. An entirely different approach in modeling of infectious diseases is to let the average number of secondary cases infected by the primary case in an otherwise susceptible population be fixed. In this situation, the transmission risk will take the form

$$p = \frac{c}{n-1}, \tag{1}$$

where c is some constant.

These two different approaches can be said to form two extreme cases regarding group size dependence, separated by a continuous array of intermediate cases. An expression for the transmission risk that covers the whole scale is

$$p = c_1 + \frac{c_2}{n-1}.$$

By letting $c_2 = 0$, we get the first case of group size independence, and by letting $c_1 = 0$, we get expression (1). In the next section, we will find reasonable estimates for c_1 and c_2 based on some real data.

To summarise both seasonal variation and group size dependence, we can write the final expression for the weekly transmission risk as

$$p(w, n) = \lambda_w \left(c_1 + \frac{c_2}{n-1} \right).$$

Let us now turn to the second transition from carrier to infected. When a susceptible individual becomes colonised through transmission of bacteria from another carrier in the population, it takes some *incubation time* T before symptoms appear if they ever do. After this period we assume that the individual develop symptoms with probability q_s and, in that case, becomes infected.

Let us again consider the state of the process at time t. Then we know that there are $C(t, t-T)$ carriers in the population that were colonised T weeks ago. (We assume throughout that T is an integer.) Each one of these will now develop symptoms with probability q_s independently of each other, which means that we can describe the total number of carriers that will become infected as

$$Y(t) \sim \text{Bin}(C(t, t-T); q_s).$$

The two last transitions will be considered simultaneously, since they are quite similar in nature, i.e. they both describe the process of recovery of carriers and infected individuals. For most infectious diseases, the immune system adapts to an ongoing infection by producing antibodies, which will eventually eradicate the infectious agent. This means that

the probability that an infected individual will recover during some specified time interval will increase with time since infection. This does not seem to be the case neither for pneumococcal carriage nor for infection. In the study by Ekdahl et al. [18], data suggest that the periods of carriage and of infection both can be regarded as exponentially distributed random variables, which implies that the probability of recovery is constant in time.

Therefore, let us introduce the probabilities

$$\begin{aligned} q_c &= \text{the probability of recovery within a week for a carrier,} \\ q_i &= \text{the probability of recovery within a week for an infected individual.} \end{aligned}$$

Again, assuming that all individuals recover independently of each other implies that

$$\begin{aligned} Z(t,s) &\sim \text{Bin}(C(t,s);q_c), \qquad s \neq t-T, \\ Z(t,t-T) &\sim \text{Bin}(C(t,t-T)-Y(t);q_c), \\ U(t) &\sim \text{Bin}(I(t);q_i). \end{aligned}$$

Note that we exclude the carriers that develop symptoms at time t, since they obviously cannot recover at the same time.

Having specified the numbers of individuals that make transitions from one state to another during the time interval from t to $t+1$, we can summarise this as follows

$$\begin{aligned} S(t+1) &= S(t) - X(t) + Z(t) + U(t), \\ C(t+1,s) &= C(t,s) - Z(t,s), \qquad s \neq t+1, s \neq t-T, \\ C(t+1,t+1) &= X(t), \\ C(t+1,t-T) &= C(t,t-T) - Y(t) - Z(t,t-T), \\ I(t+1) &= I(t) + Y(t) - U(t). \end{aligned}$$

Finally, to complete the model, we have to specify the absence mechanism, i.e. whether a child is present in the day-care group in a certain week or absent for any other reason than pneumococcal infection. To simplify things, we assume that if an individual is absent, then he or she remains absent for the entire week. Furthermore, we also assume that the events that an individual is absent in two consecutive weeks are independent.

By introducing the probability

$$q_a(w) = \text{the probability that an individual will be absent on week number } w,$$

we get the following expressions

$$\begin{aligned} S^*(t) &\sim \text{Bin}(S(t);1-q_a(w)), \\ C^*(t) &\sim \text{Bin}(C(t);1-q_a(w)). \end{aligned}$$

Note that we allow the absence probability to vary over the year. Individuals are probably more likely to be absent in the winter, when infections usually are more frequent than at other times of the year.

Month	Weeks	Absence probability
January	1–4	0.14
February	5–8	0.13
March	9–12	0.10
April	13–17	0.11
May	18–21	0.09
June	22–25	0.06
July	26–30	1.00
August	31–34	0.06
September	35–39	0.08
October	40–43	0.08
November	44–47	0.13
December	48–52	0.13

Table 1: *Probability for a child to be absent from day-care during one week in different months during the year.*

3 Parameter values

In the model, described in the previous section, there are several parameter values that have to be determined before the model can be of any practical use. Most of these will be estimated from real data, mainly from the SSPIP, but for some, where good data are not available, numerical values will simply be assumed from experience.

We will first consider the absence probabilities $q_a(w)$ $(w = 1, 2, \ldots, 52)$. In the study by Söderström and Blennow [19], monthly absence figures are reported for 599 children in day-care in Haninge south of Stockholm during 1995. We have simply used these values for the corresponding weeks, which are summarised in Table 1. Furthermore, we have assumed that day-care is closed in July, because of summer vacation. Hence, all absence probabilities for those weeks are equal to one.

Next, we will focus our attention on the recovery probabilities q_c and q_i. In the SSPIP, 935 cases of PRP carriage were followed during the period 1995-98 by weekly bacterial cultures until two consecutive negative cultures were obtained. Of these, 494 were carriers and 296 were infected children below the age of seven, the usual age in day-care in Sweden. The average time of colonisation was 4.46 weeks for carriers and 6.77 weeks for infected children.

As mentioned earlier, the study by Ekdahl et al. [18] suggests that time of carriage is exponentially distributed, which was the reason for introducing constant recovery probabilities in the model. Since we are measuring time of carriage in whole weeks, the data sets can be considered as samples of geometric rather than exponential distributions with parameters q_c and q_i, respectively. Since the mean of the geometric distribution is the

reciprocal of its parameter, we get the estimates

$$\begin{aligned} \hat{q}_c &= 1/4.46 = 0.224, \\ \hat{q}_i &= 1/6.77 = 0.148. \end{aligned}$$

The incubation time T and the probability of symptoms q_s are two quantities that are very difficult to measure. It is generally believed that the incubation time for pneumococcal infection lies somewhere in the interval 1-3 weeks. We will therefore consider the two extreme cases $T = 1$ and $T = 3$ in the following analysis. For the probability of symptoms, we have assumed that $q_s = 0.15$ [8].

Next, we will consider the seasonal variation of the weekly transmission risk $p(w, n)$, i.e. the parameters λ_w $(w = 1, 2, \ldots, 52)$. Between January 1st 1995 and June 30th 1998, there were 1545 cases of PRP carriage registered in the SSPIP. The idea is to use these data to relate, for each week of the year, the number of new cases to the number of present carriers in the population of Malmöhus county. This should give us a rather good picture of the overall transmission intensity.

Unfortunately, these data are most likely heavily biased. Some of the cases were registered because they developed symptoms and, consequently sought medical care. However, most of them were found through contact tracing around an index case. We have no idea how many asymptomatic carriers there actually were during this period. Anyway, since the number of symptomatic cases is not biased and since it seems reasonable to assume that the number of carriers is at least proportional to the number of infected individuals, it suffices to consider only the symptomatic cases.

Of the 1545 cases mentioned above, there were 642 symptomatic cases. For all of these, the date of the first positive bacterial culture was registered and, for 354 of them, also the time of carriage. The date of the first positive culture is of course not identical to the date of colonisation. First there is the incubation time T and then an additional delay of approximately one week before the treating physician suspects that the infection is caused by PRP and takes a sample for cultivation. Depending on the value of T, the date of colonisation hence precedes the date of the first positive culture by approximately 2-4 weeks. For simplicity, we assume the total delay to be equal to 3.

From these data, we obtain estimates of the number of symptomatic carriers for each week during the study period. We calculate the ratio of new cases and symptomatic carriers for each week during the calendar year. The result is shown in Figure 1.

As we can see, the estimates vary quite heavily between different weeks. The reason is probably due more to uncertainty of the estimates rather than true underlying seasonal variation. Therefore, we have also calculated a smoothed version, using kernel smoothing with a Gaussian kernel, which is also included in Figure 1. Hopefully, this smoothed curve is a better representation of the real seasonal variation and that is the one we have used in the model.

Finally, we will obtain estimates of the parameters c_1 and c_2 in the expression for the weekly transmission risk $p(w, n)$. We will again base the estimation procedure on a data set from the SSPIP. One important intervention measure within this project is to

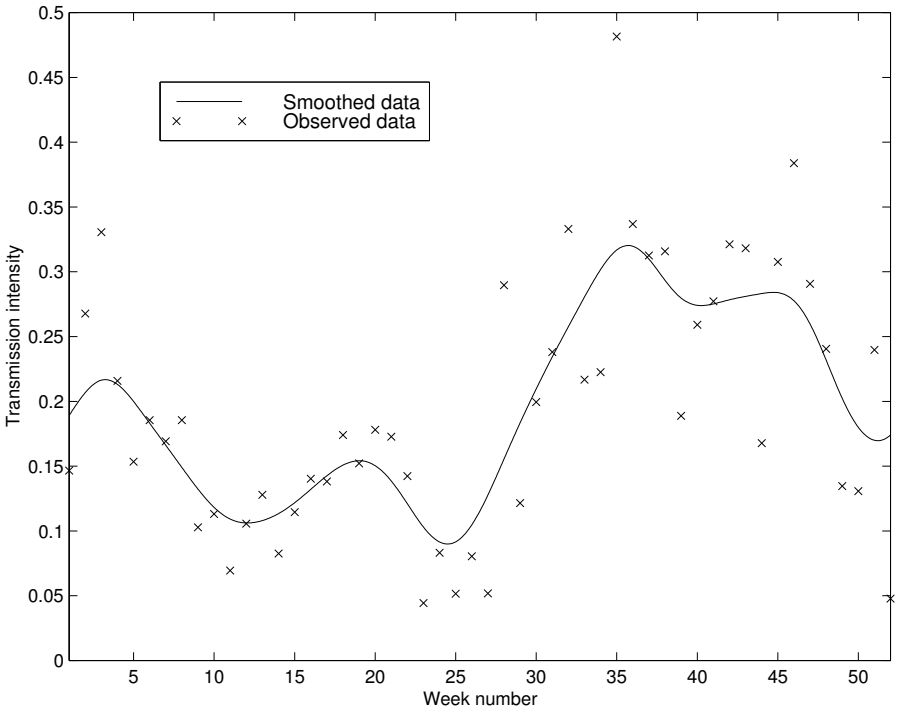

Figure 1: *Estimated transmission intensity of pneumococci in 68 day-care groups in southern Sweden (observed and smoothed data from the SSPIP study).*

identify and isolate as many asymptomatic carriers as possible through contact tracing of the symptomatic index cases. If an index case attends day-care, samples from all children in his or her day-care group are cultured and those who are found to be PRP-positive are isolated. From December 1996 to June 1998 were 98 day-care groups cultured in this way and the results from 68 of them contain sufficient information to be useful for our purposes. For these, we know the culture dates for the index case and the whole group, the size of the group and the number of positive cultures.

By using the model with all parameter values obtained so far, we can, in principle, calculate the probability distribution of the number of positive cultures as a function of c_1 and c_2 for each of the 68 day-care groups in the sample. By introducing the random variables

$$\eta_j = \text{the (random) number of positive cultures in group } j, \qquad j = 1, 2, \ldots, 68,$$

we can express the distribution of group j as

$$p_j(x; c_1, c_2) = P_{(c_1,c_2)}(\eta_j = x), \qquad x = 0, 1, 2, \ldots, n_j - 1,$$

where n_j is the size of group j and P is the usual probability measure. Then we can write

the *likelihood function* as

$$L(c_1, c_2) = \prod_{j=1}^{68} p_j(x_j; c_1, c_2),$$

where x_j is the observed number of positive cultures in group j. The maximum likelihood estimates of c_1 and c_2 are obtained by maximising $L(c_1, c_2)$.

In practice, we encounter some problems that have to be considered first. In order to use the model properly, we have to know which week the primary case was colonised, which is not the case in this situation. We know when the index case was found to carry PRP, which we can use to estimate when he or she was colonised. However, the index case is not necessarily identical to the primary case. Since the probability of developing symptoms is assumed to be only 0.15, this is actually quite unlikely. Still, for simplicity, we have assumed that this is the case. We will probably miss the actual week of introduction of bacteria into the group by a couple of weeks, but, hopefully, this will not affect the result too much. Again, we assume that the delay between the time of first colonisation in the group and the time of first positive culture for the index case is 3 weeks.

Unfortunately, the probability distributions of the variables η_j are impossible to derive explicitly, especially because of the seasonal variation in the transmission risk. Instead, we will estimate the likelihood function through computer simulations of the outbreaks in the 68 day-care groups for different values of c_1 and c_2 and then find the maximum by interpolation.

Each simulation was made in the following way. We start by letting one individual be colonised three weeks before the first positive culture of the index case was found and all the other individuals be susceptible. Then we simulate the outbreak according to the model in the previous section until some individual develops symptoms. If the spread ceases without this happening, we just start a new simulation. From this moment, we continue two more weeks; one week because of the delay before the index case is cultured and one week because of the delay before the whole group is cultivated, and observe the number of carriers in the group.

To get reliable estimates, each group is simulated 10 000 times and $p_j(x_j; c_1, c_2)$ is estimated by the proportion of the number of times where x_j is observed. By taking the product of all these estimates, we then get an overall estimate of $L(c_1, c_2)$. In this way, we obtain estimates of the likelihood function for a set of points (c_1, c_2) close to the maximum and finally find the point where the maximum is attained through two-dimensional quadratic interpolation. The result becomes

$$\begin{aligned} \hat{c}_1 &= 0.055, \\ \hat{c}_2 &= 0.48 \end{aligned}$$

for $T = 1$ week and

$$\begin{aligned} \hat{c}_1 &= 0.037, \\ \hat{c}_2 &= 0.40 \end{aligned}$$

for $T = 3$ weeks.

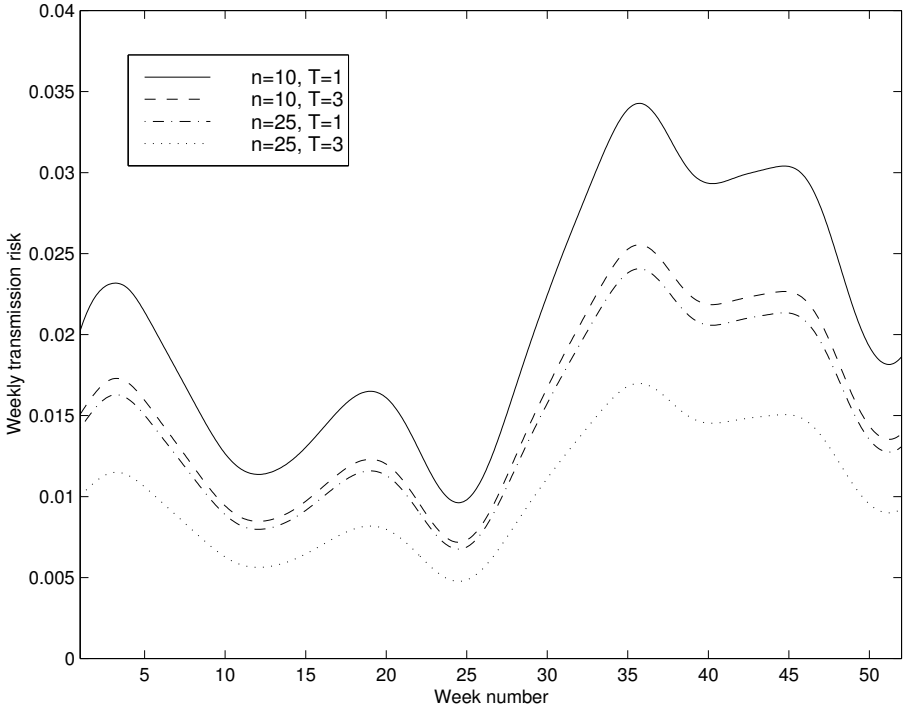

Figure 2: *Estimated weekly transmission risk of pneumococci in a day-care group (n = group size, T = incubation time).*

4 Results

The estimated weekly transmission risk

$$\hat{p}(w,n) = \hat{\lambda}_w \left(\hat{c}_1 + \frac{\hat{c}_2}{n-1} \right)$$

is shown in Figure 2 for group sizes $n = 10$ and $n = 25$ and incubation times $T = 1$ and $T = 3$. We see again that the seasonal variation is very pronounced. The maximal weekly transmission risk, which is attained in the first week of September, is 3.5 times larger than the minimal, which is attained in the middle of June. The group size dependence is also apparent but not quite that strong.

The weekly transmission risk illustrates well the infectivity of the disease on the individual level. On the other hand, it is not a good measure of the overall transmission rate within the whole population. To get a better picture of that, we will consider the *basic reproduction rate* R_0 defined as

> $R_0 =$ the average number of individuals that a carrier will transmit bacteria to during his or her whole period of carriage in an otherwise completely susceptible population.

(See e.g. [20, 17].) Intuitively, this means that the average number of secondary cases is approximately R_0, the average number of tertiary cases R_0^2 and so on. Hence, there is only a possibility of a large outbreak if $R_0 > 1$.

To derive an expression for R_0, we will use the formula

$$R_0 = \beta \cdot \kappa \cdot D$$

from Giesecke 1994, where

$$\begin{aligned} \beta &= \text{the transmission risk per contact,} \\ \kappa &= \text{the average number of contacts per time unit,} \\ D &= \text{the average duration of the infectious period.} \end{aligned}$$

In this situation, we immediately get that $\beta = p(w, n)$ and, since we assume that all individuals have close contacts, that $\kappa = n - 1$.

The average duration of the infectious period, which here means the average number of weeks a colonised individual remains carrier until he or she either recovers or develops symptoms, is a bit more complicated. For $T = 1$, we can regard the duration of the time of carriage of a randomly chosen individual as a positive integer valued random variable ξ with probability mass function

$$\begin{aligned} P(\xi = 1) &= q_c + q_s, \\ P(\xi = k) &= (1 - q_c)^{k-2}(1 - q_c - q_s)q_c, \qquad k = 2, 3, \ldots \end{aligned}$$

The average or expected value can be obtained through rather straightforward but tedious calculations as

$$D = \mathrm{E}[\xi] = \sum_{k=1}^{\infty} kP(\xi = k) = \cdots = \frac{1 - q_s}{q_c} \approx \frac{1 - q_s}{\hat{q}_c} = \frac{1 - 0.15}{0.224} = 3.79.$$

For $T = 3$, the corresponding probability mass function becomes

$$\begin{aligned} P(\xi = 1) &= q_c, \\ P(\xi = 2) &= (1 - q_c)q_c, \\ P(\xi = 3) &= (1 - q_c)^2(q_c + q_s), \\ P(\xi = k) &= (1 - q_c)^{k-2}(1 - q_c - q_s)q_c, \qquad k = 4, 5, \ldots \end{aligned}$$

with the average

$$D = \sum_{k=1}^{\infty} kP(\xi = k) = \cdots = \frac{1 - q_s(1 - q_c)^2}{q_c} \approx \frac{1 - q_s(1 - \hat{q}_c)^2}{\hat{q}_c} = \frac{1 - 0.15 \cdot 0.776^2}{0.224} = 4.06.$$

To summarise, the estimated basic reproduction rate can now be expressed as

$$\hat{R}_0 = \hat{\lambda}_w \left(\hat{c}_1 + \frac{\hat{c}_2}{n - 1} \right) (n - 1)\hat{D} = ((n - 1)\hat{c}_1 + \hat{c}_2)\hat{\lambda}_w \hat{D},$$

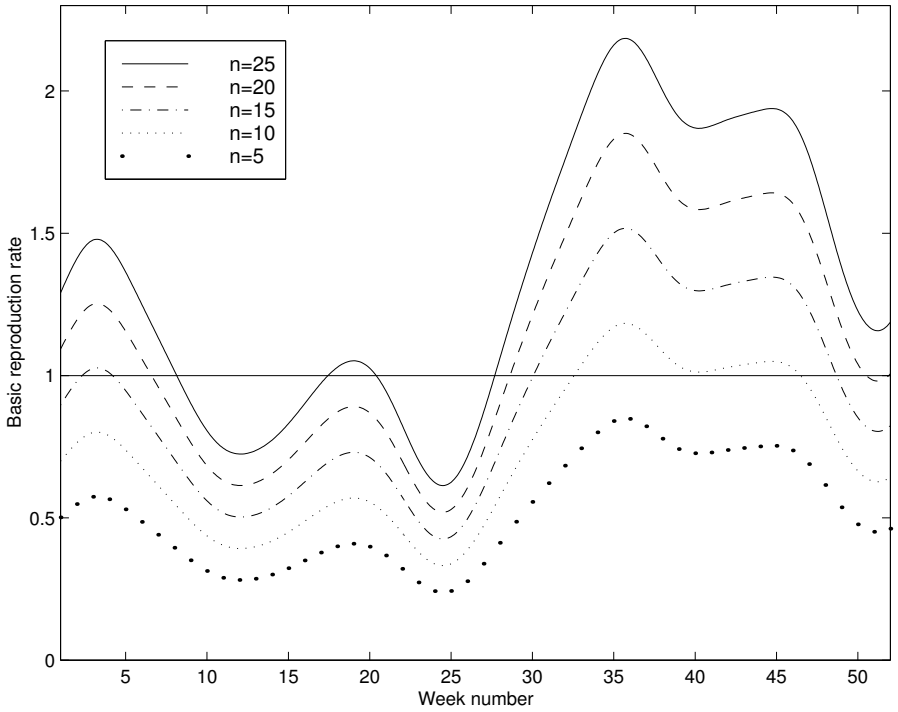

Figure 3: *Estimated basic reproduction rate for one week incubation time.*

which is shown in Figure 3 for $T = 1$ and in Figure 4 for $T = 3$ for five different group sizes ($n = 5, 10, 15, 20, 25$). Note that although the weekly transmission risk is lower in larger groups, the basic reproduction rate is higher, which is a consequence of the fact that an individual has more potentially infectious contacts in a bigger population.

As we recall, one of the main objectives of this study was to use the model to investigate the efficiency of the intervention measures within the SSPIP. This will be done by using the model to simulate outbreaks with and without any intervention and compare the results. One commonly used measure of the impact of an epidemic process is the *severity*, which usually is defined as the sum of all infectious periods over all infected individuals in the population. In some sense, we can interpret this as the total "quantity" of infection that a susceptible individual is exposed to during the whole course of an outbreak. In our situation, it is the carriers that are crucial in this respect, since we assume that all infected individuals are absent from day-care as long as they have symptoms.

With the notation introduced in Section 2, we can define the severity without intervention as

$$s_0(w_0, n) = \sum_{t=0}^{\infty} C(t),$$

and with intervention as

$$s_1(w_0, n) = \sum_{t=0}^{\theta+2} C(t),$$

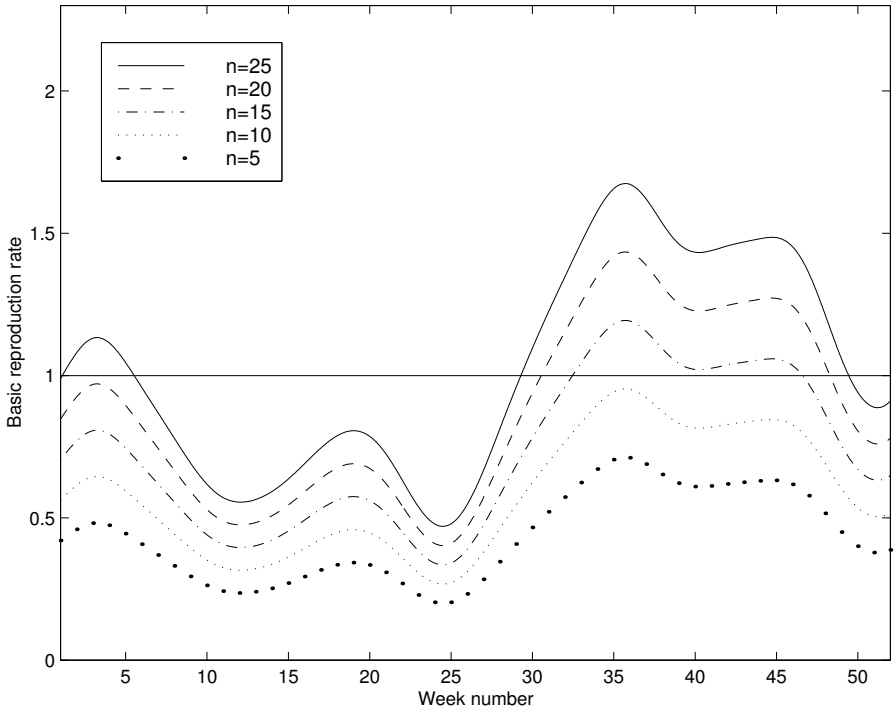

Figure 4: *Estimated basic reproduction rate for three weeks incubation time.*

where θ is the time when the first individual develops symptoms, formally

$$\theta = \min\{t : I(t) > 0\}.$$

As mentioned earlier, we assume a delay of two weeks from the week when the first individual develops symptoms until the intervention measures are carried out. If no individual develops symptoms before the process dies out, we let $\theta = \infty$.

Since time of the year and group size obviously play important roles in the spreading dynamics, we emphasise this by letting the severity be a function of the week when the bacteria is introduced w_0 and the group size n. In the analysis, we will consider eleven different weeks ($w_0 = 1, 5, 9, 13, 18, 22, 31, 35, 40, 44, 48$), which correspond to the first week of each month except July when we assume that day-care is closed, five different group sizes ($n = 5, 10, 15, 20, 25$) and two different incubation times ($T = 1, 3$). In total, this comprises 110 different combinations of which we made 10 000 simulations each to estimate the severity with and without intervention. The result is shown in Tables 2 and 3.

5 Discussion

The model, presented in Section 2, is particularly designed to take two important aspects regarding spread of pneumococci into account, namely *seasonal variation* and the *size* of

	Without intervention					With intervention				
Group size	**5**	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**
January	4.9	6.0	7.2	9.1	12.0	4.5	5.1	5.9	6.3	7.0
February	4.7	5.7	6.5	8.6	11.0	4.4	5.1	5.6	6.2	6.8
March	4.3	5.2	6.0	6.9	9.4	4.3	4.7	5.0	5.6	5.9
April	4.4	4.9	5.9	7.2	9.8	4.1	4.6	5.0	5.5	5.8
May	4.4	5.2	6.4	8.3	11.4	4.3	4.7	5.2	5.7	6.3
June	4.2	5.1	6.4	8.0	11.9	4.2	4.5	4.9	5.5	5.9
August	6.0	8.9	13.6	22.8	37.8	5.2	6.3	7.6	9.0	10.1
September	6.3	9.4	14.3	22.9	37.2	5.6	6.7	8.2	9.6	11.1
October	6.0	8.0	11.5	17.6	26.2	5.3	6.2	7.4	8.8	9.9
November	5.7	7.2	9.7	13.5	19.6	5.1	6.0	6.9	7.9	8.9
December	5.2	6.6	7.8	11.1	14.8	4.8	5.4	6.3	6.8	8.0

Table 2: *Average severity without and with intervention for one week incubation time.*

day-care groups.

As Figure 1 suggests, the most important of these is probably the seasonal variation, i.e. the way the transmission risk varies over the calendar year. We see that the most crucial time of the year is the autumn with a maximum estimated transmission intensity in week number 36, which corresponds to the first week in September. This could be a consequence of the fact that some children may have been colonised during summer vacation with new strains of bacteria against which other children have no acquired immunity. Then it decreases somewhat steadily with a small peak in January (after Christmas vacation) and a bigger peak in May until it reaches minimum in the middle of June. Another cause behind this pattern could be that children in ordinary day-care tend to be more indoors in bad or cold weather, thereby having closer contact with each other. Also, the presence of other respiratory illnesses, which are more common in autumn and winter, may facilitate the spread of pneumococci. However, the quite pronounced peak in late spring does not seem to have any intuitive explanation.

Group size also seems to be an important factor, although not quite as significant as seasonal variation. Nevertheless, when they interact, i.e. when considering large groups in the autumn, the result is rather severe. As mentioned in the previous section, whether or not the basic reproduction rate R_0 is bigger than one can change the behaviour of an epidemic process substantially. If $R_0 < 1$, a potential outbreak always dies out quickly without affecting more than a few individuals. On the other hand, as soon as $R_0 > 1$, there is a possibility of an outbreak so that bacteria may spread to several individuals and remain within a day-care group for a long time. According to Figures 3 and 4, groups with more than ten children seem to be at risk during most of the autumn and bigger groups during the winter as well, if the incubation time is only one week. However, for the other alternative of three weeks incubation time, it seems that group sizes up to around fifteen

	Without intervention					With intervention				
Group size	**5**	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**
January	5.2	5.9	7.2	8.1	10.1	5.1	5.6	6.3	7.2	8.3
February	5.1	5.7	6.7	7.7	9.2	5.0	5.6	6.2	7.0	7.9
March	4.8	5.4	5.8	6.7	7.6	4.8	5.2	5.5	6.2	6.8
April	4.7	5.2	5.7	6.6	7.8	4.6	4.9	5.5	5.9	6.6
May	4.9	5.6	6.5	7.6	9.0	4.8	5.1	5.7	6.4	7.2
June	4.7	5.3	6.1	7.2	9.2	4.6	5.1	5.6	6.1	6.7
August	6.5	8.5	11.9	17.6	26.6	6.0	7.4	9.4	11.2	13.7
September	6.8	9.3	12.8	17.8	27.2	6.4	7.8	9.7	12.3	15.2
October	6.3	8.0	10.3	14.0	19.7	5.9	7.2	8.8	10.7	12.5
November	6.1	7.3	9.2	11.5	15.0	5.8	6.7	8.0	9.4	11.0
December	5.7	6.5	7.8	9.3	11.7	5.4	6.1	6.9	8.1	9.3

Table 3: *Average severity without and with intervention for three weeks incubation time.*

children are not at risk. It should be pointed out though that the estimated effect of group size, i.e. the estimates $\hat{c}_1$ and $\hat{c}_2$, is rather uncertain, since it is based on the small sample of 68 day-care groups.

Not having reliable knowledge nor a good estimate of the incubation time for pneumococci infection is naturally a big limitation and, as we saw in the previous section, the two likely extreme cases $T = 1$ and $T = 3$ produce somewhat different results. Figure 2 shows that the estimated transmission risk decreases with longer incubation time. This effect is, to some extent, compensated by a longer duration of carriage when considering the basic reproduction rate, especially for small groups (Figures 3 and 4). An attempt was made to assess which of the two cases that was most consistent with data from the 68 day-care groups, but no strong evidence was found to support either case.

One thing that should be noted is the comparatively low weekly transmission risk presented in Figure 2. As we see, it typically lies in the range 1–4 % depending on the time of the year, group size and incubation time. But this is the risk that one specific child will be colonised during one week when there is only one carrier present in the group. If we consider several other carriers and a longer time period, this risk may increase substantially.

Two other aspects, which have been shown to be significant in other studies [21, 22, 18], but have been disregarded here are age of the children and serotype of pneumococci. It seems that younger children are colonised more easily and for longer periods than older. Up until around two or three years of age, the immune system is not yet fully developed [23], but after that there does not appear to be any significant difference with respect to age [21, 7]. Therefore, a slightly more complex model was considered, where individuals were subdivided into two different types; children up to three years old and children older than three years. However, no clear distinction in parameter estimates appeared, so in order not to complicate the model more than necessary, we chose to disregard the age of the

children.

Within the SSPIP, 18 different serotypes of pneumococci were detected during the period 1995-98. The spread of one of these has been described in detail by Melander et al. [24]. Although there are studies indicating that different serotypes have different epidemiological characteristics [22, 18], available data do not imply any major diversity. Some serotypes seemed to be transmitted more easily and some were carried for a longer time but the differences were not pronounced. Again, to not introduce unnecessary complexity into the model, we made no distinction between the different serotypes.

When comparing the severity of pneumococci carriage without and with intervention in Tables 2 and 3, three interesting patterns can clearly be seen.

Firstly, intervention is much more efficient in the second half of the year than in the first. For instance, the average severity in August to December in a group of 25 children for the case $T = 1$ is reduced from 27.1 weeks to 9.6, which means a relative efficiency of 65 %. The corresponding reduction in January to June is only from 10.9 to 6.3 weeks, a relative efficiency of 43 %. This is naturally a consequence of the higher basic reproduction rate in the autumn. When the intervention is made, the outbreak is most likely in its initial stage and is hence stopped before the bacteria can establish in the group. However, it may take some time before an outbreak is detected. First there is an incubation time of 1–3 weeks, then only 15 % of colonised individuals develop symptoms and finally it takes approximately two weeks before it is found that the index case carries PRP. Anyway, since spread is actually rather slow because of the low weekly transmission risk, this apparently long delay does not seem to be too serious.

Secondly, intervention is more efficient in large groups than in small. In August, the average severity in a group of 20 children is reduced from 22.8 to 9.0 weeks for $T = 1$, while in a group of just 10 children it is reduced from 8.9 to 6.3 weeks. Again, the higher basic reproduction rate in bigger groups is the main factor behind this. For $n = 20$, the estimated basic reproduction rate in August is $\widehat{R}_0 = 1.3$ but for $n = 10$ only $\widehat{R}_0 = 0.8$.

Thirdly, intervention is more efficient for shorter incubation times. As we saw above, the average severity in August in a group of 20 children was reduced from 22.8 to 9.0 for $T = 1$. The corresponding reduction for $T = 3$ is from 17.6 to 11.2. It is quite easy to understand why the reduction decreases, since a longer incubation time implies a longer delay before the intervention is implemented. It might seem a bit contra-intuitive that the severity without intervention is smaller for longer incubation times. A longer incubation time means a longer average duration of carriage, i.e. a colonised individual will spend more time in the day-care group before possibly developing symptoms and hence spread the bacteria more efficiently. But this argument only holds under the condition that the transmission risk is the same for both incubation times, which is not the case here. We have obtained separate estimates of the weekly transmission risk for $T = 1$ and for $T = 3$ (see Figure 2) and since it turned out to be smaller for longer incubation times, the severity becomes smaller as well.

To summarise, it seems that the intervention measures do have a significant impact on the spread of pneumococci within day-care groups, at least in large groups and in the second half of the year. Whether this is enough to stop overall spread in society is hard

to say, it depends of course on how many of all infectious contacts take place within day-care groups. Of all 642 index cases, i.e. those PRP-carriers that developed symptoms and therefore sought medical care, detected during 1995–98 within the SSPIP, approximately 300 were children who attended day-care. Although this does not mean that almost half of all transmissions take place within day-care groups, it still suggests that all day-care centres are an important environment for the spread of pneumococci.

The main objective of this study was to construct a model for the spread of pneumococci within one single day-care group and to use this to evaluate the kind of intervention measures that are taken within the SSPIP. The reason why we restricted ourselves to such a small population, besides being an important sub-population, was mainly that it appears to be quite homogeneous and that we had access to reliable data upon which to base estimates of unknown parameters. A natural next step would be to extend the model to the level of day-care centres. It seems rather obvious that quite a lot of transmissions take place between day-care groups as well as within. Although the transmission risk between children who attend different groups at a day-care centre probably is much smaller than between children in the same group, there are most likely sufficiently many infectious contacts between groups to play an important role in the spreading process. Such an analysis would require a more complex model of a more heterogeneous population and also more extensive data in order to obtain good estimates of transmission risks between groups, and thus falls outside the scope of this study.

References

[1] Austrian R. Some aspects on the pneumococcal carrier state. *J. Antimicrob. Chemother.* 1986; **18**: 35–45.

[2] Huebner RE, Wasas A, Mushi A, Mazhani L, Klugman K. Nasopharyngeal carriage and antimicrobial resistance in isolates of streptococcus pneumoniae and haemophilus influenzae type b in children under 5 years of age in Botswana. *Int. J. Infect. Dis.* 1998; **3**: 18–25.

[3] Kalin M. Bacteremic pneumococcal pneumonia: Value of nasopharynx culture and examination of washed sputum specimens. *Eur. J. Clin. Microbiol.* 1982; **1**: 394–396.

[4] Söderström M. *Respiratory tract infections in children. Aspects of etiology, incidence and recurrent infections.* Thesis, Studentlitteratur: Lund, Sweden, 1990.

[5] Strangert K. Respiratory illness in preschool children with different forms of day care. *Pediatrics* 1976; **57**: 191–196.

[6] Ståhlberg MR. The influence of form of day-care on occurence of acute respiratory tract infections among young children. *Acta Paediatr. Scand.* 1980; **282**: 1–87.

[7] Wald E, Guerra N, Byers C. Frequency and severity of infections in day care: Three year follow-up. *J. Pediatr.* 1991; **118**: 509–514.

[8] Gray BM, Converse GM, Huhta N, Johnston Jr RB, Pichichero ME, Schiffman G, Dillon Jr HC. Epidemiologic studies of *Streptococcus pneumoniae* in infants: antibody response to nasopharyngeal carriage of types 3, 19 and 23. *J. Infect. Dis.* 1981; **144**:312–318.

[9] Appelbaum PC. Antimicrobial resistance in *Strptococcus pneumoniae*: an overview. *Clin. Infect. Dis.* 1992; **15**: 77–83.

[10] Baquero F, Garcia-Rodriguez JA, Garcia de Lomas J, Aguilar L. Antimicrobial resistance of 1,113 *Streptococcus pneumoniae* isolates from patients with respiratory tract infections in Spain: Results of a 1-year (1996–1997) multicenter surveillance study. *Antimicrob. Agents Chemother.* 1999; **43**: 357–359.

[11] Marton A. Pneumococcal antimicrobial resistance: the problem in Hungary. *Clin. Infect. Dis.* 1992; **15**: 106–111.

[12] Olsson-Liljequist B, Burman LG, Kallings I. Antibiotic susceptibility of upper respiratory tract pathogens in Sweden: a seven year follow-up study including loracarbef. Swedish Respiratory Tract Study Group. *Scand. J. Infect. Dis.* 1992; **24**: 485-493.

[13] Ekdahl K, Kamme C. Increasing resistance to penicillin in *Streptococcus pneumoniae* in southern Sweden. *Scand. J. Infect. Dis.* 1994; **26**: 301–305.

[14] Forsgren A, Walder M. Antimicrobial susceptibility of bacterial isolates in south Sweden including a 13-year follow-up study of some respiratory tract pathogens. *APMIS* 1994; **102**: 227–235.

[15] Ekdahl K, Hansson HB, Mölstad S, Söderström M, Walder M, Persson K. Limiting the spread of penicillin-resistant *Streptococcus pneumoniae*: experiences from the South Swedish Pneumococcal Intervention Project. *Microbial Drug Res.* 1998; **4**: 99–105.

[16] Bailey NTJ. *The mathematical theory of infectious diseases and its applications.* Griffin: London, 1975.

[17] Anderson RM, May RM. *Infectious diseases of humans: Dynamics and control.* Oxford Univ. Press: Oxford, 1991.

[18] Ekdahl K, Ahlinder I, Hansson HB, Melander E, Mölstad S, Söderström M, Persson K. Duration of nasopharyngeal carriage of penicillin-resistant *Streptococcus pneumoniae*: experiences from the South Swedish Pneumococcal Intervention Project. *Clin. Infect. Dis.* 1997; **25**: 1113–1117.

[19] Söderström M, Blennow M. Barn på utedagis hade lägre sjukfrånvaro. (in Swedish) *Läkartidningen* 1998; **15**: 1670–1672.

[20] Giesecke J. *Modern infectious disease epidemiology.* Edward Arnold: London, 1994.

[21] Harsten G, Prellner K, Heldrup J, Kalm O, Kornfält R. Acute respiratory tract infection in children. A three year follow-up from birth. *Acta Pediatr. Scand.* 1990; **79**: 402–409.

[22] Smith T, Lehman D, Montgomery J, Gratten M, Riley ID, Alpers MP. Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. *Epidemiol. Infect.* 1993; **111**: 27–39.

[23] Andersson U, Bird A-G, Britton S, Palacios R. Humoral and cellular immunity in humans studied at the cell level to two years of age. *Immunol. Rev.* 1981; **57**: 5–38.

[24] Melander E, Ekdahl K, Hansson HB, Kamme C, Laurell M, Nilsson P, Persson K, Söderström M, Mölstad S. Introduction and clonal spread of penicillin- and trimethoprim/sulfamethoxazole-resistant *Streptococcus pneumoniae*, serotype 9V, in southern Sweden. *Microbial Drug Res.* 1998; **4**: 71–78.