Mathematical Statistics
Stockholm University

Occurrence of invasive pneumococcal disease and number of excess cases due to influenza

Katarzyna Grabowska

Examensarbete 2005:6

Postal address:

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

Internet:

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

Mathematical Statistics
Stockholm University
Examensarbete **2005:6**,
http://www.math.su.se/matstat

Occurrence of invasive pneumococcal disease and number of excess cases due to influenza

Katarzyna Grabowska*

September 2005

Abstract

Influenza is associated with seasonal outbreaks, often with a high rate of morbidity and mortality. It is also known to be a cause of significant amount secondary bacterial infections. *Streptococcus pneumoniae* is the main pathogen causing secondary bacterial pneumonia after influenza and subsequently, influenza could participate in obtaining Invasive Pneumococcal Disease (IPD). In this study, we aim to investigate the relation between influenza and IPD and estimate the yearly excess of IPD cases due to influenza. For this purpose, we use influenza periods as an indicator for influenza activity as a risk factor in subsequent analysis. The statistical modelling has been made in two modes. First, two negative binomial regression models have been constructed, and by estimating the contribution of influenza in the models, excess number of IPD cases has been calculated. Secondly, an "influenza free" baseline was constructed, and differences in IPD data and baseline were used to estimate a yearly additional number of IPD cases due to influenza. Both modes were calculated using zero to four weeks lag time. The analysis shows a yearly increase of 72–118 IPD cases due to influenza, which corresponds to 12–24 % per influenza season. Also, a lag time of one to three weeks appears to be of significant importance in the relation between IPD and influenza.

*Postal address: Smittskyddsinstitutet, SE–171 82 Solna, Sweden. E-mail: kasia.grabowska@smi.ki.se. Supervisor: Åke Svensson.

Contents

Chapter 1

Introduction

1.1 Background

Influenza, with its annual epidemics, is the infection associated with highest mortality in the developed world. In the United States, an average of more than 18,000 annual deaths are due to influenza, and the influenza epidemics cause almost 50 000 hospitalisations in influenza and pneumonia annually among elderly people [1].

Streptococcus pneumoniae is the most important bacteria causing secondary pneumonia after influenza [2][3][4]. Invasive pneumococcal disease (IPD) is associated with a high case fatality rate, despite modern intensive care [5][6]. Known risk factors for IPD are age (the very old and the very young), male sex, and underlying debilitating conditions [7]. In Sweden, the overall incidence of IPD is 15 per 100,000 and year, and in the above 65 years age group the incidence may be as high as 40 to 50 per 100,000 [5].

Both IPD and influenza have distinct seasonal patterns, with winter peaks [8][9][10]. Besides this annual seasonal pattern with peaks of influenza incidence during the temperate winter season, there are year-to-year variations both in intensity and timing of occurrence [10][11][12]. Schwartzman et al was the first to document a temporal association between influenza and IPD in the early 1970s [13]. The same finding has later been described in studies from Scotland [14], the Netherlands [15], and the United States [4][16]. In several early, hospital-based studies, pneumococcal pneumonia has been shown to be a complication of clinical diagnosis of influenza. [21][22][23]

The aim of this study was to more in detail model the association

between influenza and IPD using Swedish surveillance data. The main aim of this project is to evaluate if the incidence of IPD is increased during the influenza season. A secondary aim, if the results show a relationship between the IPD and influenza, is to establish the increase in IPD incidence, related to influenza.

Poisson regression has often been used in epidemiological studies when an outcome variable (e.g. number of IPD cases) is a rare occurrence, and Poisson regression has recently been applied in studies investigating the association between influenza and mortality [17][18]. However, one of the main characteristics of the Poisson model is that its variance equals its mean. In other words, if a Poisson model is fitted to data with a variance greater than its mean (overdispersion), the variance will be underestimated. To overcome this limitation, in this thesis, three negative binomial models are constructed, which give the same estimation of a mean value as a Poisson model, but include overdispersion in the variance [19].

1.2 Data

IPD data

Since 1994, all invasive pneumococcal isolates obtained in Sweden (one per patient) have been reported to the Swedish Institute for Infectious Diseases Control (SMI). An isolate is defined as invasive if it is retrieved from blood, cerebrospinal fluid or other normally sterile sites. All such isolates, reported to the SMI from 1 January 1994 to 28 March 2004 (n=12 010) were included in the study. Fifty percent of the isolates were from elderly persons (65 years of age of older), 43% were from adults (20-64 years of age), and 7% of the isolates were from children and teenagers. In the analysis, the date of culture to SMI was used. Since this date is not always specified, the samples with absent date of culture were omitted (n=373). The IPD data set is illustrated in Figure 1.1.

Influenza data

In Sweden, the influenza surveillance is based on weekly reports (Monday to Sunday of each individual week) on the number of influenza cases diagnosed by the local laboratories to SMI. Serology reports are not included in these

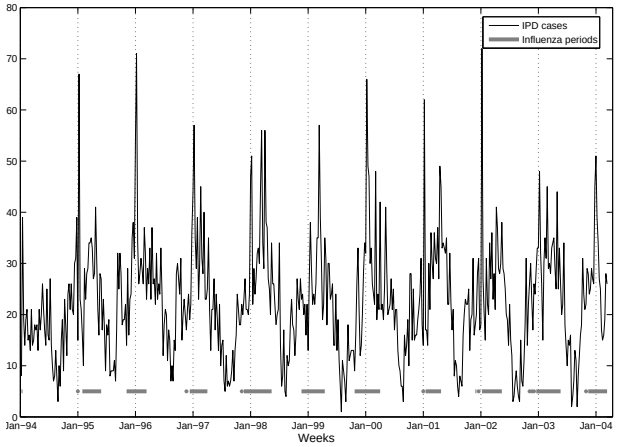

Figure 1.1: Number of IPD diagnosis per week from Jan 1994 to March 2004. Also, the horizontal lines indicate the period of time when influenza was present.

reports. From 1994 to 2001, the registration of influenza occurred from week 43 to week 16. From 2001/2002, the influenza-reporting season was extended from week 40 to 20. For the influenza reports, we used the same time period (January 1994 to March 2004) as for the IPD isolates (n=10,498). The age distribution of influenza cases varied between individual years. Data on age of influenza cases are available from the season 1998/1999, and shows that an average of 47% of the cases were elderly (above 65 years) and 39 % adults (20-64 years) during the period 1998-2004.

We defined influenza activity as presence of laboratory reports of influenza cases. As the laboratory cases mainly originate from patients with severe disease in need of hospital care, this data might not reflect the true influenza activity among the general population. Since the season 1999-2000 an additional component has been added to the Swedish influenza surveillance system, based on reports of the number of patients with influenza-like illness from about 120 sentinel physicians in outpatient care. During the time when both systems have been running in parallel, the time periods for influenza activity have coincided [17][20], supporting that the laboratory surveillance system adequately mirrors community influenza activity despite the selected

population it is based on. This consistency does also indicate that even though the criteria for culturing might have changed during the observation period, it has not affected the definition of the annual influenza period by the laboratory reporting.

In the models, an indicator for influenza was used (one if influenza is present and zero if absent). This approach was found more suitable since, the influenza data are based on laboratory reports, and as mentioned above, mainly reflects the hospitalized patients and not the population as total.

1.3 Disposition

This report starts with an introduction into the nature of IPD and the challenges that arises when this type of data is investigated. In the introduction, the aim and the structure of the data is specified. In Chapter 2 the analysis and description of the applied models are made. The details of the computational part of the analysis models are presented in Chapter 3. This chapter also contains information about the characteristics of each model and its ability to establish the relationship between the IPD data and influenza. The discussion and conclusions are specified in Chapter 4. Finally, all derivations and additional results related to this report, are presented in Appendix A.

Having the readers in mind, in order to smooth the progress of studying this report, the studied data is applied on the theory throughout the report. Hence, the general theory is not singled out but recurringly applied on the specific problem. Also, in this thesis the terminology for the natural logarithm is ”log” and not ”ln”.

Chapter 2

Regression models

2.1 Poisson regression

Poisson Regression model is suitable, when an event occurs (as independent observations) a number of times, Y, during a given period of time. Here, observations $\mathbf{y} = y_1, y_2, ...y_n$, are number of observed IPD cases during week i for ten whole years, and 10 weeks of year 11 (hence, $n = 10 \cdot 52 + 10 = 530$). Also, one characteristic for a Poisson variable is that the mean $\mathrm{E}[Y]$ and variance $\mathrm{Var}(Y)$ are equal:

$$\mathrm{E}[Y] = \mathrm{Var}[Y] = \mu$$

The mean value μ, can be estimated from the data and affects several explanatory variables. We denote each explanatory variable as x_{ij} where the index i denote week $i = 1, ..., 530$, and j explanatory variable $j = 1, ..., r$. Hence a Poisson Regression Model with r variables can generally be written as:

$$\log(\mathrm{E}[Y_i|x_{ij}]) = \log(\mu_i) = \beta_0 + \Sigma\beta_j x_{ij}$$

where β_0 is the intercept and β_j are parameters for explanatory variable j. The investigated explanatory variable of main interest in this report is influenza.

To be able to capture the cyclic, recurring seasonality in that is observed in the data, see Figure 1.1, the regression model needs to include a term that affected the amplitude A and lag φ for each season. Note that the lag represents a fraction of a year at the time t, $t = 1, ..., 52$).

$$A\cos\left(\frac{2\pi}{52}\cdot t-\varphi\right)$$

To include the term in the model it was rewritten during the simulation as:

$$A\cos\left(\frac{2\pi}{52}\cdot t-\varphi\right)=a\cos\left(\frac{2\pi}{52}\cdot t\right)+b\sin\left(\frac{2\pi}{52}\cdot t\right)$$

from which A and φ can be computed as $A=\sqrt{a^2+b^2}$, $\varphi=\arctan(b/a)$, see Appendix A.2 for details. A general model that captures the the yearly trend, seasonal variation and influence of influenza can then be presented as:

$$\log(\mu_i)=\beta_{\text{sin}}\sin\left(\frac{2\pi}{52}t_i\right)+\beta_{\text{cos}}\cos\left(\frac{2\pi}{52}t_i\right)+\beta_{\text{interc}}+\beta_{\text{infl}}x_{i0}+\beta_{\text{year}}x_{i1} \quad (2.1)$$

where β_{interc} is the intercept, x_{i0} is an indicator variable for influenza (0 if no influenza occurred week i or 1 if influenza was present week i and x_{i1} denotes the year. β_{year} is the year parameter, β_{infl} describes presence of influenza and β_{sin} and β_{cos} represents the periodicity parameters. t_i represent week i. The structure of the data set is presented in Table 2.1.

index (i)	influenza (x_{i0})	year (x_{i1})	week (t_i)	IPD (y_i)
1	1	1994	5	19
2	1	1994	6	21
3	0	1994	7	15
⋮	⋮	⋮	⋮	⋮
52	0	1994	52	15
53	0	1995	1	67
54	0	1995	2	23
⋮	⋮	⋮	⋮	⋮
529	0	2004	9	19
530	0	2004	10	28

Table 2.1: Structure of data.

There are some limitations of this model that should be pointed out:

- the yearly trend is constant and therefore does not allow for variation in amplitude from year to year.
- since y_i belong to a Poisson distribution, the variance equals the mean in the model.

Finally, one constraint that is to be kept in mind in the Poisson and Negative binomial regression modelling is that the response variable y_i, is considered to be independent from week i to $i+1$ or, in other words: $Y_1, Y_2, ..., Y_{530}$ are assumed to be independent.

2.2 Negative binomial regression

One of the disadvantages of the previous model was the equality between the mean and the variance. There are some arguments that promote a higher variance (rather than equal to its mean) in this data set. One factor is a random variation of number of IPD.

It is possible to check if a model need to include higher variance (than the mean), so called overdispersion[1]. This can be done by calculating the Pearson χ^2 statistics:

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i}$$

Now, dispersion is estimated by X^2/n. (If there is no overdispersion, the statistics, X^2 and n are approximately equal, the ratio is than ≈ 1 and the data is Poisson distributed.) Calculations of Pearson χ^2 statistics indicate that overdispersion is present, see Section 3, and that the variance is greater than the mean. Hence the Poisson assumption is not valid.

A model that is equivalent to the Poisson model but allows for a greater variance, overdispersion, is a Negative binomial model. In this model, the assumption of equal expected value and variance $\mathrm{E}[\mu] = \mathrm{Var}(\mu)$ is replaced by an additional assumption, that the intensity parameter μ, is Gamma-distributed (see Appendix A.1 for details). so that:

$$Y_i \sim \ \mathrm{Po}(\mu_i), \quad \text{where} \ \ i = 1, 2, ..., 530$$

and

$$\mu_i \sim \Gamma(\alpha_i, \delta)$$

[1]Models where the variance is greater than its mean are called overdispersed

The mean value $\mathrm{E}[\mu_i]$ and variance, $\mathrm{Var}(\mu_i)$ is then:

$$\begin{aligned} \mathrm{E}[\mu_i] &= \frac{\alpha_i}{\delta} \\ \mathrm{Var}(\mu_i) &= \frac{\alpha_i}{\delta^2} \end{aligned}$$

Now, the expected value and variance of Y_i is:

$$\mathrm{E}[Y_i] = \frac{\alpha_i}{\delta} \tag{2.2}$$

$$\begin{aligned} \mathrm{Var}[Y_i] &= \frac{\alpha_i}{\delta} + \frac{\alpha_i}{\delta^2} \\ &= \frac{\alpha_i}{\delta}\left(1 + \frac{1}{\delta}\right) \end{aligned} \tag{2.3}$$

Derivations in Equ.2.2 and 2.3 can be found in [19]. In the Poisson regression model the variance is equal to its mean, Here when comparing the $\mathrm{E}[Y_i]$ in Equ.2.2 and $\mathrm{Var}[Y_i]$ in Equ.2.3 the difference is the factor $(1 + 1/\delta)$. This factor measures the dispersion and is called the dispersion parameter.

In the Negative binomial regression model, μ is Gamma-distributed, hence, the modeled response variable Y_i is:

$$\log\left[\mathrm{E}(Y_i)\right] = \log\left(\frac{\alpha_i}{\delta}\right) = \log(\alpha_i) - \log(\delta) \tag{2.4}$$

Subsequently $\log(\alpha_i)$ and $\log(\delta)$ will represent the response variable y_i in a negative binomial model, presented in Sections 2.2 and 2.2.

Model 1

The Negative binomial regression model can now be presented as:

$$\begin{aligned} \log(\alpha_i) &= \log(\delta) + \beta_{\sin}\sin\left(\frac{2\pi}{52}t_i\right) \\ &+ \beta_{\cos}\cos\left(\frac{2\pi}{52}t_i\right) + +\beta_{\mathrm{interc}} + \beta_{\mathrm{inf}}x_{i0} + \beta_{\mathrm{year}}x_{i1} \end{aligned} \tag{2.5}$$

where:

$$\mu_i \sim \Gamma(\alpha_i, \delta)$$

Model 2

In the previous negative binomial model, Equ.2.5, the yearly trend is estimated by one parameter, β_{year}. By dividing the β_{year} parameter into 11 parameters (one for each year), the model will be specified to:

$$\begin{aligned} \log(\alpha_i) &= \beta_{\cos}\cos\left(\frac{2\pi}{52}t\right) + \beta_{\sin}\sin\left(\frac{2\pi}{52}t\right) \\ &+ \beta_{\text{interc}} + \beta_{\text{inf}}x_{i0} + \sum_{j=1}^{11}\beta_{\text{year}j}x_{ij} \\ \text{where} & \\ \mu_i &\sim \Gamma(\alpha_i, \delta) \end{aligned}$$

Baseline

The main purpose of this report was to estimate if a presence of influenza affects the number of registered IPD. One way of studying this discrepancy is to compare the data (when influenza is present and with an index variable for influenza) to a model without influenza (baseline). In order to estimate the baseline, a data set of weeks when no influenza was reported was constructed. The response variable was then:

$$Y_i = \begin{cases} y_i & \text{if number of influenza} = 0 \text{ week } i \\ \text{missing value} & \text{if number of influenza} > 0 \text{ week } i \end{cases}$$

This baseline was constructed for one season, where for each week, all values where no influenza occurred was modelled. Also, since influenza is not monitored during summer, only weeks when influenza was monitored was used.

Here, $y_{i,k}$ is Poisson distributed, $y_{i,k} \sim \text{Po}(\mu_{i,k})$ and as before $\mu_{i,k}$ is Gamma distributed. Hence a Negative binomial regression is used to estimate the baseline:

$$\begin{aligned} \log(\alpha_i) &= \beta_{\cos} \cos\left(\frac{2\pi}{52} t_{ik}\right) + \beta_{\sin} \sin\left(\frac{2\pi}{52} t_{ik}\right) + \beta_{\text{interc}} \\ \mu_{i,k} &\sim \Gamma(\alpha_i, \delta) \end{aligned}$$

where $k = 1, 2, ..., K_i$ where K_i varies between 1 and 10 (depending on number of weeks with no influenza during week i). For example: During week 1, influenza was registered for all years except two: 1995 and 2001. Hence i=1 (the observed week) and K_i=2 since we only observed two years of eleven, when no influenza occurred. Then, since $\mu_{i,k} \sim \Gamma(\alpha_i, \delta)$ the estimate of α_1 is based on two data points.

2.3 Model fitting

For all calculations, STATA 8.0 was used. The Negative binomial regression models were calculated using function 'nbreg'. Data, available for analysis was from week 1 1994 to week 10 2004. However, in order to be able to compare the models with different lags (in other words, have the same number of data for each lag), week 1 - 4 in 1994 was omitted in the analysis.

For both models, the question of interest was if the presence of influenza affects the number of IPD cases. Our hypothesis is that an arbitrary case of IPD could occur due to a previous influenza infection. In order to investigate more in detail if a predefined lag time between influenza infection and IPD disease promotes the IPD occurrence, both models were tested in four modes: no lag, 1, 2, 3 and 4 weeks lag. In order to establish if a Negative binomial regression model should be used instead of Poisson, i.e. to establish if overdispersion was present, Pearson, statistics were calculated for each of the evaluated models. See appendix A.3.1 for details.

To evaluate which model was preferred, the log-likelihood was calculated for each model.

Chapter 3

Estimation of number of excess IPD cases

In this Chapter, the results of the models described in Section 2.2 are presented. Initially, to verify that overdispersion was present (so that the Poisson model could be excluded from further analysis), the Pearson χ^2 statistic was calculated using the formulae presented in Equ. 2.2. The results are presented in table A.1in Appendix and indicate that $\chi^2/n > 1$, hence that the dispersion parameter should be included.

The analysis of the IPD data and influenza was made using two approaches:

1. Estimating number of excess cases due to the influenza parameter in the models presented in Section 2.2 with four different lags.

2. Calculating mean difference of cases from the baseline, using the same lag where the strongest significance was found in the models calculated using approach 1.

Results of estimates are presented in Table 3.1 and 3.2. Also, Model 1 and Model 2 (with 3 weeks lag) are illustrated in Figure 3.1.

3.1 Results of models

To further estimate the impact of the influenza parameter, the models with significant parameter, were recalculated without the term $\hat{\beta}_{\text{inf}}$. Setting the

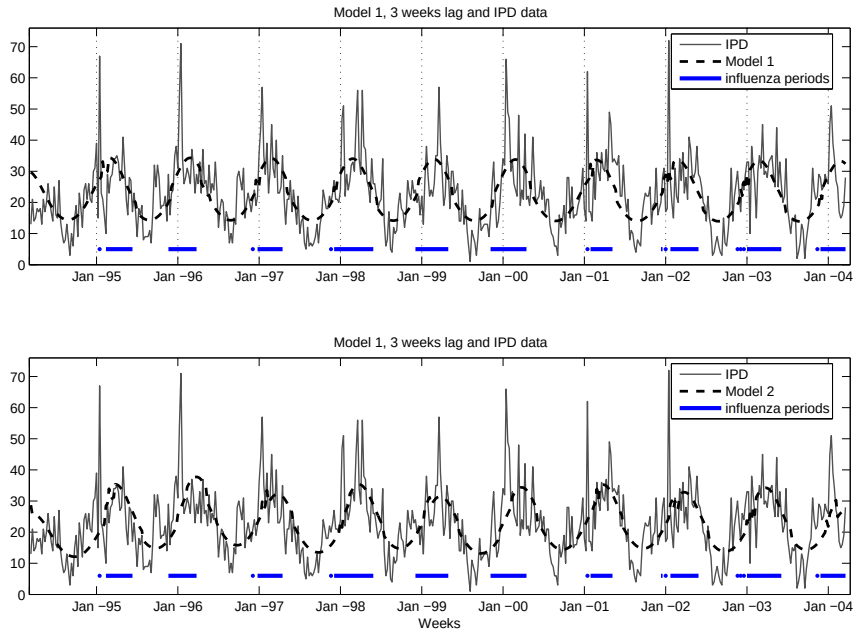

Figure 3.1: Model 1 and Model 2, both calculated with 3 weeks lag. Also, IPD data and influenza periods are illustrated.

Model 1	Influenza parameter and its 95% c.i.	P-value	Log-likelihood
No lag	0.08 (-0.01 - 0.18)	0.090	-1843.0
1 week lag	0.15 (0.06 - 0.25)	0.001	-1839.1
2 weeks lag	0.14 (0.04 - 0.24)	0.004	-1840.2
3 weeks lag	0.16 (0.06 - 0.25)	0.001	-1839.0
4 weeks lag	0.14 (0.05 - 0.24)	0.003	-1840.1

Table 3.1: Results of Approach 1, for Model 1

Model 2	Influenza parameter and its 95% c.i.	P-value	Log-likelihood
No lag	0.03 (-0.07 - 0.14)	0.55	-1831.0
1 week lag	0.12 (0.02 - 0.23)	0.025	-1828.6
2 weeks lag	0.10 (-0.01 - 0.20)	0.07	-1829.4
3 weeks lag	0.13 (0.02 - 0.23)	0.015	-1828.1
4 weeks lag	0.11 (0.01 - 0.21)	0.03	-1828.8

Table 3.2: Results of Approach 1, for Model 2

term $\hat{\beta}_{\text{inf}}$ to zero is equivalent to calculating the model with $x_{0i} = 0$. This could estimate number of IPD cases that should have occurred if no influenza was present or if influenza had no effect. The overall mean difference of the models and its 95% Confidence Intervals (c.i.) for each week was then calculated. Weekly number of excess IPD cases are illustrated in Figure 3.2.

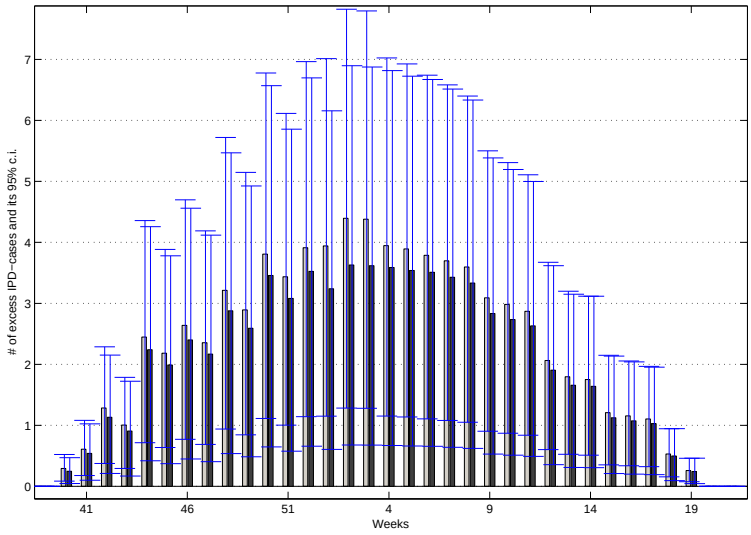

Figure 3.2: Differences in number of IPD cases, calculated by Model 1(light gray bars) and Model 2 (dark grey bars), calculated with 3 weeks lag. Also, errorbars illustrate 95% confidence intervals.

3.2 Results of baseline

In this analysis, the difference in IPD-incidence was studied, in weeks when influenza occurred (influenza weeks) compared to weeks when it was not present (influenza-free weeks) during the annual influenza registration season. The baseline was created in order to establish a mean level of IPD-incidence when influenza was absent. The baseline was modelled from all influenza-free weeks when influenza was monitored (winter season) during the 11-year study period; however, during 2003 influenza occurred in all 13 studied weeks, hence, that year did not contribute to the estimation of baseline.

In order to estimate number of excess cases of differences between IPD and baseline was calculated. This was made by summarizing total number of weekly differences, in periods when influenza was monitored and present (registered). The total sum of differences used in calculation corresponded to a 10.23 years period. (10 whole years from Jan 1994 to dec 2003 and 12 weeks of year 2004.) Hence in order to calculate number of yearly cases, the total sum was divided by 10.23.

We choose a 1 - 4 weeks lag in the calculations of baseline, since previous calculations (estimates of influenza as an index variable in the models) showed association between influenza and IPD with significant p-value, using those lags, see Table 3.1 and 3.2.

3.2.1 Bootstrap calculation of 95% confidence intervals.

In order to estimate 95% confidence intervals (c.i.) for the baseline, a data set containing differences between baseline and IPD data from weeks when influenza was present was constructed. From those differences (a total of 217 values) values of sums were bootstrapped, by 10000 simulations. When sorted, value 250 corresponded to 2.5% quantile and value 9750 to 97.5% quantile of the 95% c.i. Finally, in order to estimate the yearly 95% intervals, calculated 2.5% and 97.5% values were divided by 10.23.

Number of IPD during influenza free weeks (that occurred while monitoring influenza) and baseline estimation for lags 1 to 4 weeks are presented in Table 3.3. Also, baseline with 3 weeks lag is illustrated in Figure 3.3.

	Number of cases	% per season
Baseline (1 week lag)	150 (121 - 181)	24% (19 - 29%)
Baseline (2 week lag)	157 (127 - 188)	25% (20 - 30%)
Baseline (3 week lag)	152 (122 - 182)	24% (20 - 29%)
Baseline (4 week lag)	154 (123 - 185)	25% (20 - 30%)

Table 3.3: Number of excess IPD cases due to baseline calculations.

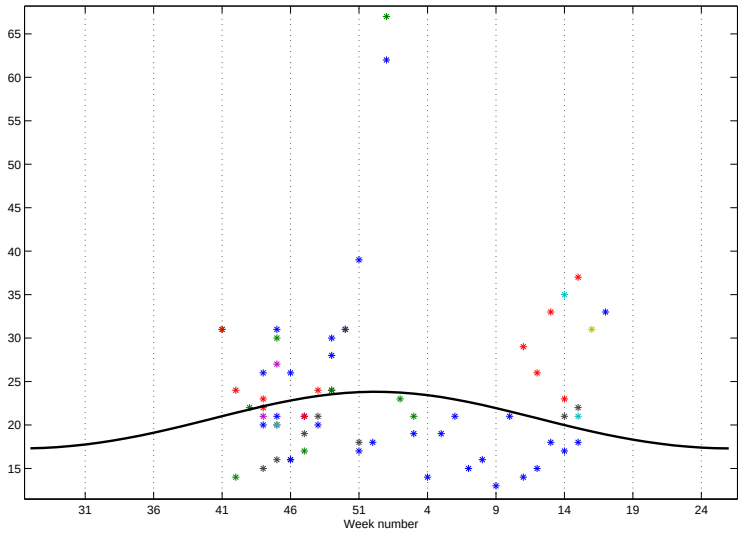

Figure 3.3: Baseline, calculated with 3 weeks lag. Also, number of IPD-cases during influenza free weeks (in time period when influenza was registered) are presented.

3.3 Number of estimated excess cases of IPD

Number of excess cases of IPD, estimated as described in Section 3.1 and 3.2 is presented in Table 3.4.

	Number of cases	% per season
Model 1 (3 weeks lag)	81 (24 - 243)	13% (4 - 24%)
Model 2 (3 weeks lag)	72 (14 - 138)	12% (2 - 23%)
Baseline (3 weeks lag)	152 (122 - 182)	24% (20 - 29%)

Table 3.4: Number of excess IPD cases

Chapter 4

Discussion

In this report, two approaches are used to study the impact of influenza on IPD: the first one estimates impact of an influenza parameter in the models, and the second estimates the mean difference of IPD cases from an influenza-free baseline. Both methods used in this study indicated an increase in number of IPD cases due to influenza. Since both take different factors into account, it is difficult to decide which one of them should be preferred. Hence, conclusions should be made taking both methods into account. The first approach, when the effect of the influenza parameter is estimated, uses influenza data as an indicator variable, which most likely underestimates the impact of influenza during the winter season. Hence, it was preferred to more likely get an underestimation of the true effect than the other way around. This could probably explain why the results in this approach give lower estimates of number of IPD cases related to influenza, than the baseline approach.

Furthermore, in this method, the importance of the influenza parameter in the model and its effect on subsequent IPD cases is studied. The result shows that one to four weeks lag puts forward the significance of influenza indicator variable in both models. Also, as it can be seen in Table 3.1 and 3.2, for both models that the size of estimate $\hat{\beta_{\text{inf}}}$ increase with the size of the lag and peak at 3 weeks lag, indicating presence of delay when taking into account an effect of influenza on IPD data. However, since influenza data is transformed to be an indicator variable of influenza, there can be no inference made about the relative changes in number of IPD by increasing or decreasing the number of influenza by one unit. The second approach, where a baseline is constructed, is based on a subset of the IPD data set, in

other words the baseline is constructed of all weeks when influenza is absent. The validity of the baseline varies over the season, with lowest variance in the spring and fall and largest variation during winter (since in this time of year, influenza is often present, hence, very few data points are used in the estimation). Since influenza is not monitored during summer, there is no data from this period used in the estimation of the models.

In order to detect a case of IPD in the surveillance system, a sample must be collected and sent to the laboratory for culturing. In the analysis, the date of culture was used as the onset date. Since onset of IPD disease is often of acute nature, it is realistic to assume that the culturing is made closely to the onset of disease.

As mentioned in previous studies [24], an increase or decrease in incidence can be observed due to altered routines in culturing. However, no active interventions or routine changes in the culturing tradition of IVP have been executed during 1994-2004.

As for IPD, laboratory tests for influenza are normally performed in an acute phase of the disease. We can therefore assume that the time estimates for occurrence of disease are correct, with no major delays between detection and report of cases.

The results may be biased only by factors that systematically change the temporal patterns (other than seasonal, annual or long-term trends) of either influenza or IPD diagnoses. One possibility of bias is therefore changes in health-care seeking behavior and patient-sampling due to holiday periods. Therefore, a dummy variable for holiday season was added into the model, with little effect on our results or conclusions.

Since Model 1 and Model 2 are not of hierarchical structure, it is difficult to interpret if one of them fits better to the data than the other. However, since for each model, the same data is used for the different lag calculations, the log-likelihoods can be compared within the models. For both model 1 and 2, the largest log-likelihood is obtained by using 3 weeks lag indicating the best fit to the data. On the other hand, the log-likelihood differ (within the models) very little, see Table 3.1 and 3.2.

This study confirms the association between the two diseases even after taking into account seasonal variation, and also shows that the strength of this association is highly seasonal with a peak excess of IPD morbidity due to influenza in January, see Figure 3.2. However, the data sets (influenza data and IPD data) have not been compared on an individual level, hence it is not possible to establish if single individuals are present in both data

bases.

Finally, the results, considering the association between influenza and IPD morbidity, coincide, despite the fact that two different approaches are used to define number of excess IPD cases. Furthermore, the detection of a lag between the influenza and IPD morbidity falls out to be an important component in forecasting amount of IPD cases, hence public health measures against influenza and IPD are preferably considered together.

Bibliography

[1] CDC. Updated recommendations from the Advisory Committee on Immunization Practices in response to delays in supply of influenza vaccine for the 2000-01 season. MMWR 2000;49:888-92.

[2] Treanor JJ. Orthomyxoviridae: influenza virus. In: Mandell GL, Douglas DR, Bennett JE, Dolin R (eds). Principles and Practice of Infectious Diseases, 5th Ed. New York: Churchill Livingston, 2000; 1834-5.

[3] Robinson KA, Baughman W, Rothrock G, et al. Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998. JAMA 2001;285:1729-35.

[4] O'Brien KL, Walters MI, Sellman J, Quinlisk P, Regnery H, Schwartz B, et al. Severe pneumococcal pneumonia in previously healthy children: the role of preceding influenza infection. Clin Infect Dis 2000;30:784-9.

[5] Ekdahl K, Mårtensson A, Kamme C. Bacteremic pneumococcal infections in Southern Sweden 1981-1996: trends in incidence, mortality, age-distribution, serogroups and penicillin-resistance Scand J Infect Dis 1998; 30:257-62

[6] Örtqvist Å, Kalin M, Julander I, Mufson MA. Deaths in bacteremic pneumococcal pneumonia. A comparison of two populations– Huntington, WVa, and Stockholm, Sweden. Chest 1993; 103:710-6.

[7] Musher DM. Infections caused by Streptococcus pneumoniae: clinical spectrum, pathogenesis, immunity, and treatment. Clin Infect Dis 14: 801-807, 1992

[8] Dowell, S.F., et al., Seasonal patterns of invasive pneumococcal disease. Emerging Infectious Diseases, 2003. 9(5): p. 573-9.

[9] Goddard NL, Kyncl J, Watson JM. Appropriateness of thresholds currently used to describe influenza activity in England. Commun Dis Public Health 2003; 6(3): 238-45

[10] Monto, A.S., Occurrence of respiratory virus: time, place and person. Pediatric Infectious Disease Journal, 2004. 23(1 Suppl): p. S58-64.

[11] Lina, B., et al., Surveillance of community-acquired viral infections due to respiratory viruses in Rhone-Alpes (France) during winter 1994 to 1995. Journal of Clinical Microbiology, 1996. 34(12): p. 3007-11.

[12] Nguyen-Van-Tam, J.S., et al., Excess hospital admissions for pneumonia and influenza in persons > or = 65 years associated with influenza epidemics in three English health districts: 1987-95. Epidemiology & Infection, 2001. 126(1): p. 71-9.

[13] Schwarzmann, S.W., et al., Bacterial pneumonia during the Hong Kong influenza epidemic of 1968-1969. Archives of Internal Medicine, 1971. 127(6): p. 1037-41.

[14] Kyaw, M.H., et al., Incidence of invasive pneumococcal disease in Scotland, 1988-99. Epidemiology & Infection, 2002. 128(2): p. 139-47.

[15] de Neeling, A.J., et al., Temporary increase in incidence of invasive infection due to Streptococcus pneumoniae in The Netherlands. Clinical Infectious Diseases, 1999. 29(6): p. 1579-80.

[16] Kim PE, Musher DM, Glezen WP, Rodriguez-Barradas M, Nahm W, Wright C. Association of invasive pneumococcal disease with season, atmospheric conditions, air pollution and the isolation of respiratory viruses. Clinic Infect Dis 1996; 22: 100-6.

[17] Thompson, W.W., et al., Mortality associated with influenza and respiratory syncytial virus in the United States. JAMA, 2003. 289(2): p. 179-86.

[18] Gemmell, I., et al., Seasonal variation in mortality in Scotland. International Journal of Epidemiology, 2000. 29(2): p. 274-9.

[19] Linbäck J. A model for analysing temporal and spatial patterns of infectious diseases with an application to reported campylobacterinfections,

Dep. of mathematical statistics. 2003, Stockholm University: Stockholm

[20] The National Influenza reference Centre, Swedish Institute for Infectious Disease Control (SMI) . Annual Report September 2003 -August 2004. Solna : SMI; 2004, [http://www.smittskyddsinstitutet.se/upload/PDF-filer/WHO-influensarapport-03-04.pdf]

[21] C.M. Martin et al., Asian influenza A in Boston, 1957-1958.II. Severe staphylococcal pneumoniacomplicating influenza. AMA Arch Intern Med, 1959 103(4):p532-42

[22] R.G. Petersdorf et al., Pulmonary infectionscomplicating Asian Influenza. AMA Arch Intern Med, 1959. 103(2): p. 262-72

[23] C. Jarstrand and G. Tunevall, The influence of bacterial superinfectionon the clinical caurse of influenza. Studies from teh epidemic in Stockholmduring the winters 1969-70 and 1971-72. Scand J Infect Dis, 1975 7(4): p.243-7

[24] Konradsen, H.B. and M.S. Kaltoft, Invasive pneumococcal infections in Denmark from 1995 to 1999: epidemiology, serotypes, and resistance. Clinical & Diagnostic Laboratory Immunology, 2002. 9(2): p. 358-65.

Appendix A

Appendix

A.1 Definition of Gamma distribution

In the literature, a various number of re-writings of the definitions for the Gamma distribution are presented. In this paper, the Gamma distribution is defined as:

$$\Gamma(a, b) \quad \text{where} \quad a > 0, \quad b > 0$$

with a density function:

$$f(x) = \frac{1}{\Gamma(a)} x^{a-1} b^a e^{-bx}$$

Furthermore, since μ_i is $\Gamma(\alpha_i, \delta)$ distributed, the mean value $\mathrm{E}[\mu_i]$ and variance, $\mathrm{Var}(\mu_i)$ is then:

$$\begin{aligned} \mathrm{E}[\mu_i] &= \frac{\alpha_i}{\delta} \\ \mathrm{Var}(\mu_i) &= \frac{\alpha_i}{\delta^2} \end{aligned}$$

A.2 Seasonal factors

Show

$$A\cos(kx - \varphi) = a\cos(kx) + b\sin(kx) \tag{A.1}$$

where:

$$\begin{aligned} A &= \text{amplitude} \\ \varphi &= \text{lag} \\ k &= \frac{2\pi}{52} \\ a \text{ and } b &= \text{constants} \end{aligned}$$

Knowing that:

$$\cos(s-t) = \cos(s)\cos(t) + \sin(s)\sin(t)$$

we can rewrite the left side of equation A.1:

$$\begin{aligned} A\left(\cos(kx-\varphi)\right) &= A\left(\cos(kx)\cos(\varphi) + \sin(kx)\sin(\varphi)\right) \\ &= \underbrace{A\cos(\varphi)}_{a}\cos(kx) + \underbrace{A\sin(\varphi)}_{b}\sin(kx) \end{aligned}$$

The relation between a, b and φ is:

$$\varphi = \arctan\left(\frac{b}{a}\right)$$

since:

$$\frac{b}{a} = \frac{A\sin(\varphi)}{A\cos(\varphi)} = \tan(\varphi)$$

Furthermore:

$$\sqrt{a^2+b^2} = \sqrt{A^2\cos^2(\varphi) + A^2\sin^2(\varphi)} = A\underbrace{\sqrt{\cos^2(\varphi)+\sin^2(\varphi)}}_{=1} = A$$

This show that equality in Equ.A.1 is valid when $\sqrt{a^2+b^2} = A$ and $\varphi = \arctan\left(\frac{b}{a}\right)$.

A.3 Results

A.3.1 Pearson χ^2 statistic

For both models and all lags, Pearson χ^2 vas calculated. Results are presented in

A.3.2 Estimates of Parameters

	χ^2/n for Model 1	χ^2/n for Model 2
No lag	3.22	3.12
1 week lag	3.19	3.10
2 weeks lag	3.23	3.13
3 weeks lag	3.19	3.09
4 weeks lag	3.21	3.11

Table A.1: Pearson χ^2 statistic

Parameter	Coef.	Std. Err.	z	$P > \|z\|$	Lower 95% C.I.	Upper 95% C.I.
$\hat{\beta}_{\text{year}}$	-0.0009	0.0056	-0.16	0.876	-0.0120	0.0102
$\hat{\beta}_{\text{inf}}$	0.1413	0.0484	2.92	0.003	0.0464	0.2361
$\hat{\beta}_{\text{sin}}$	0.2610	0.0323	8.08	0.000	0.1977	0.3243
$\hat{\beta}_{\text{cos}}$	0.2501	0.0257	9.71	0.000	0.1996	0.3006
$\hat{\beta}_{\text{interc}}$	3.0185	0.0374	80.63	0.000	2.9451	3.0918
δ	2.2248	0.2038			1.8197	2.6625

Table A.2: Results Model 1

Parameter	Coef.	Std. Err.	z	$P > \|z\|$	Lower 95% C.I.	Upper 95% C.I.
$\hat{\beta}_{\text{inf}}$	0.1264	0.0518	2.44	0.015	0.0249	0.2279
$\hat{\beta}_{\text{sin}}$	0.2768	0.0324	8.54	0.000	0.2133	0.3403
$\hat{\beta}_{\text{cos}}$	0.2495	0.02680	9.29	0.000	0.1969	0.3021
$\hat{\beta}_{\text{year1}}$	0.0611	0.1267	0.48	0.630	-0.1873	0.3095
$\hat{\beta}_{\text{year2}}$	0.2608	0.1186	2.20	0.028	0.0284	0.4932
$\hat{\beta}_{\text{year3}}$	0.3260	0.1185	2.75	0.006	0.0938	0.5583
$\hat{\beta}_{\text{year4}}$	0.1684	0.1197	1.41	0.160	-0.0662	0.4030
$\hat{\beta}_{\text{year5}}$	0.2592	0.1185	2.19	0.029	0.0261	0.4914
$\hat{\beta}_{\text{year6}}$	0.1441	0.1197	1.20	0.229	-0.0910	0.3787
$\hat{\beta}_{\text{year7}}$	0.2327	0.1194	1.95	0.051	-0.0014	0.4668
$\hat{\beta}_{\text{year8}}$	0.2576	0.1190	2.16	0.030	0.0244	0.4909
$\hat{\beta}_{\text{year9}}$	0.1866	0.1195	1.56	0.119	-0.0477	0.4208
$\hat{\beta}_{\text{year10}}$	0.2285	0.1188	1.92	0.054	-0.0043	0.4613
$\hat{\beta}_{\text{interc}}$	2.8066	0.1122	25.01	0.000	2.5867	3.0266
$\hat{\delta}$	2.0775	0.1944			1.7294	2.4958

Table A.3: Results Model 2