Mathematical Statistics
Stockholm University

Campylobacter infections in Sweden – A statistical analysis of temporal and spatial distributions of notified sporadic campylobacter infections

Johan Lindbäck
Åke Svensson

Research Report 2001:4

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 **2001:4**,
http://www.matematik.su.se/matstat

Campylobacter infections in Sweden – A statistical analysis of temporal and spatial distributions of notified sporadic campylobacter infections

Johan Lindbäck
Åke Svensson*

February 2001

Abstract

In Sweden, diagnosed campylobacter infections are notifiable by law. Each year during 1992-1999 approximately 2000 cases of indigenous cases were reported. Our prime aim is to find out if and how information of sporadic cases obtained through the notification system can be used to discern temporal and spatial patterns. The time series of reported cases shows a regular behaviour. Each year there is a high incidence period during the late summer and early autumn. Data, for each of 21 counties, are fitted to a model. The model includes parameters that describe the time of the peak, the duration of the high incidence period and its amplitude. The parameters may vary between years and counties. The computations are made using MCMC techniques. The model fits well to the observed data.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: akes@matematik.su.se. Financial support from The Bank of Sweden Tercentenary Foundation is gratefully acknowledged.

Table of contents

1 Introduction 1
2 Campylobacter, infectious agent and disease 2
  2.1 The agent 2
  2.2 The disease 2
  2.3 The spread of the disease 2
3 Basic facts of the reporting system 4
  3.1 Incidence of campylobacter infections 4
  3.2 The reporting system 5
  3.3 Age and sex 6
  3.4 Geography 7
  3.5 Seasonability 10
4 Analysis of temporal patterns 10
  4.1 The data 10
  4.2 A crude analysis of temporal and spatial distribution of cases 11
  4.3 Smoothing by MCMC-methods 13
  4.4 The Poisson assumption and overdispersion 17
5 Results 19
  5.1 The fit of the model 19
  5.2 Some results 21
6 Conclusions 23
7 Plans for future research 24
Acknowledgement 25
8 References 25
Appendices 27

1 Introduction

In Sweden, around fifty infectious diseases are notifiable by law. If a physician or a laboratory finds that a patient has one of these diseases, they have to send a report of the case to Smittskyddsinstitutet, SMI (Swedish Institute for Infectious Disease Control). The purpose is surveillance of the spread of infectious diseases, to follow if there are changes in spread patterns. Thus, a vast number of reports, with potentially interesting information regarding the occurrence of infectious diseases during long time periods are accumulated at SMI. Traditionally the number of reports is summed and basic statistics in form of time series or geographical distribution of the reported cases are published in monthly and yearly reports.

As is the case with most reporting systems there are severe problems with the quality of the reporting. Any analysis based on reported cases has to consider to what extent important features of the disease is reflected in the available data. One should hope that even if there is underreporting and biases in the reporting system some real effects are still seen such as temporal and spatial patterns. Understanding such patterns can be an essential contribution to understanding of how the different infectious diseases are spread. At best the reports can be used to derive important information about the occurrence and aetiology of the notifiable infectious diseases.

The prime aim of the present study is to find out if and how information provided through the notification system can be used. To do this we have taken reports on campylobacter infections as an example to find out which the problems and the possibilities are. Campylobacter was chosen for several reasons. It is an important disease both as regards its consequences and as regards the number of individuals that suffer from it. This means that the number of reported cases is sufficiently large to make patterns in temporal and spatial variability visible. It is also a disease, which has been subject of intense interest. Some, but far from all, is known of its aetiology and spread patterns. Our hope is that experiences gained from a study of campylobacter reports could be helpful to indicate how statistics regarding other infectious diseases can be analysed.

Campylobacter infections in humans are part of a possibly complicated system of spread of campylobacter bacteria. The bacteria can be found in nature, both in wild and domestic animals and in the environment. There is a transmission of the campylobacter bacteria in nature to humans (often through food contamination). This means that campylobacter infections correlate to the existence of bacteria in the environment and in food. Thus, it is of interest to study reports in connection with other data, which can explain the occurrence of the bacteria in nature, e.g., seasonal variations and variations related to weather conditions as well as data of presence of campylobacter bacteria in the environment and in the food chain. However, in this first report we will concentrate on describing temporal and spatial variations of reported human cases. This description will, in a later study, be used as a starting point for studies of connections with other phenomena. How this can be done is discussed in a final section of this report.

This work is a part of a larger project, supported by The Bank of Sweden Tercentenary Foundation, which aims to study and develop models for statistical analysis of infectious diseases.

2 Campylobacter, infectious agent and disease

2.1 The agent

Campylobacter infection is a bacterial disease. There are over 20 subtypes of *Campylobacter* but the main types causing gastro-intestinal symptoms in humans are *Campylobacter jejuni* and *Campylobacter coli*. The bacteria can survive 4 weeks in water at +4°C but less than 4 days at +25°C (cf. Notermans (1995) and Andersson & Gustavsson (1998)). To be able to multiply the bacteria requires a temperature of +40°C and a concentration of oxygen of at most 5 %. Therefore, the ideal place to grow is the intestines in humans and warm-blooded animals. Hence, food items are not a good place for the bacteria to grow but on the other hand, the critical infectious dose (the smallest dose of the bacteria to cause the disease) is very low.

2.2 The disease

Campylobacter infection is a zoonosis, which means that it is naturally transmitted between animals and man. The symptoms characterising the disease are diarrhoea, abdominal pain, malaise, fever, nausea and vomiting. The incubation period is usually 1 to 3 days but can vary between 1 to 10 days. The illness is acute and usually over within 2 to 5 days. Campylobacter infection causes 5 % - 14 % of diarrhoea worldwide and is an important cause of travellers' diarrhoea (cf. Chin (2000). There is no vaccine against the disease.

2.3 The spread of the disease

Even a superficial study of statistics reveals that cases occur both in large outbreaks and as sporadic cases. An outbreak occurs, when many individuals are exposed to the same source of infection, and suffer from the disease at approximately the same time. The sporadic cases involve only one or a few individuals that are infected simultaneously.

The cause of a large outbreak is often relatively easy to identify. During the period studied, one can discern three major outbreaks.

- In Kramfors (in the county Västernorrland) approximately 2 500 cases of campylobacter infections occurred during May 1994. This outbreak was caused by contaminated water (cf. Andersson et al (1994)). Of these cases, 64 were reported to SMI (cf. Smittskyddsinstitutet (1995)).

- In Mark (in the county Västra Götaland) 3 000 – 4 000 campylobacter infections occurred at the end of May 1995. The cause of the outbreak was contaminated water (cf. Bresky et al (1995)). Not more than 48 of the cases were reported to SMI (cf. Smittskyddsinstitutet (1996)).

- The third outbreak took place at a training camp for young football players in the summer of 1996. At least 123 out of 200 participants were infected after drinking unpasteurized milk. Of these cases 22 were reported to SMI (cf. Smittskyddsinstitutet (1997)).

Even if underreporting of cases in connection with large outbreaks is severe, they are still identifiable in a crude time series of reported cases. In the following, we concentrate on cases that appear to be sporadic. This means that we have removed observations from the large outbreaks in the data that we analyse.

In addition to these large outbreaks there may occur minor outbreaks involving only a few individuals. Evidently, such minor outbreaks are much more difficult to identify. The yearly reports from SMI mentions seven such minor outbreaks in 1998 (cf. SMI (1999)). These outbreaks resulted in 3 – 7 reported cases. The source of infection varied. Identified or suspected causes were food (unpasteurized milk, chicken, paella) or drinking contaminated water.

It has been discussed if campylobacter infections are communicable. The general understanding seems to be that there is a small risk of spread human to human, but that it is rather limited (cf. Chin (2000)). We have chosen not to include effects of such spread in the models and our analysis.

The substantial part of reported cases are sporadic i.e. they cannot be seen to be directly associated with other cases. As for cases during outbreaks, sporadic cases are often not reported. However, one can expect the underreporting of sporadic cases to be less severe. There exists no reliable investigation of the proportion of unreported cases in Sweden. In an English study, it was established that for each reported case of C*ampylobacter*, in a laboratory based surveillance system, there were 7.6 cases in the community (cf. Wheeler et al (1999)). We cannot assume that this number is representative for Swedish conditions due to the many differences between England and Sweden concerning both the surveillance systems and the communities.

In some of the reports from the physicians, a suspected cause of the infection is mentioned. Such causes are badly prepared chicken, chicken prepared at home, eating at restaurant, secondary infections, barbecue, contact with birds, drinking unpasteurized milk or water from mountain brooks.

There are a number of investigations of risk factors associated with sporadic cases (cf. Kapperud (1995)). Many of the studies are case-control studies. Comparing to which extent cases and healthy controls have been exposed to potential risk factors one tries to identify exposures that increase the risk of getting a campylobacter infection. A few examples of risk factors studied are travel abroad, contacts with animals and food consumption. There are of course severe difficulties in managing studies of this kind. The quality of the study relies on that sufficiently good accounts of the exposures for the cases before taking ill and similar reliable accounts of exposures for the controls can be obtained. The results can be subject to recall bias since cases and controls remember or report their true exposures with different accuracy. The danger of recall bias is even larger when the participants in the study are asked to recall if they have been eating a certain food item, and to make an evaluation of the exposure (e.g. if the chicken they have consumed was undercooked chicken or not). A well-established risk factor has been found in a series of studies from Great Britain were humans were infected due to birds (mainly magpies) pecking off the seals of milk bottles (cf. Lighton et al (1991)). Otherwise the results presented in published studies are, as can be expected, rather diffuse. Kapperud (1995) lists identified risk factors in a number of case-control studies made in different countries. The list contains travel abroad, eating chicken, handling raw chicken, eating undercooked chicken, eating chicken at barbecues, eating poultry, eating at barbecues, drinking surface water, drinking untreated water, drinking raw milk, drinking raw goat's milk, milk bottles pecked by magpies, contact with cats, presence of a puppy in the household. This broad spectrum of risks can be taken as an indication that there are several routes of transmission of campylobacter infections to humans. Due to the large seasonal variations of (reported) campylobacter infections, it is possible that different transmission routes are open at different times of the year. Even if Tauxe (1992) calculates that 50 % of the cases are attributable to consumption of poultry products, there seems to be no single risk factor that accounts for the most of the cases. Of course, it may be the case that infections have different causes in different surroundings and at different times.

3 Basic facts of the reporting system

3.1 Incidence of campylobacter infections

From 1992 to 1997 on average 5 000 cases per year of campylobacter infections were reported to SMI (table 3.1). During 1998 and 1999, there was an increase in incidence. In 1998, infections acquired in Sweden and abroad both increased as compared with the previous year. However, the number of domestic cases was at the same level as 1994 and 1995. The increase in incidence in 1999 compared with 1998 is due to an increase in cases infected abroad. The major part of the reported cases of campylobacter infections relates to persons travelling outside of Sweden. Between 31 % and 46 % of the cases each year are infected in Sweden and the rest are infected abroad

Table 3.1. Number of reported cases of campylobacter infections in Sweden by origin of infection

	Year of registration at SMI							
Place of infection	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
In Sweden	1 453	1 825	2 538	2 551	1 815	1 828	2 586	2 209
Abroad	2 998	2 590	2 764	2 821	3 131	3 266	3 816	4 796
Unknown	24	70	227	208	136	212	142	132
Total	4 475	4 485	5 529	5 580	5 082	5 306	6 544	7 137

In the following analysis, we are only considering infections that have been acquired in Sweden.

3.2 The reporting system

There are of course all kinds of quality problems associated with this compulsory notification system. Even if a disease is notifiable by law not all cases are reported properly. Any analysis of geographical and temporal patterns will rely on the precise information given in the reports and how this information is processed at SMI.

From the time an individual is infected till a report of the resulting illness ends up in the registers at SMI several steps have to be passed. First, the infected person has to go to the doctor. For diseases with mild symptoms, this can lead to a reasonable under-reporting because many infected persons will not seek medical help. Then the physician has to make the correct diagnosis, which may be confirmed by a laboratory test. Subsequently a report must be filled in, signed, and sent to SMI. At SMI, the reports are entered into a database. Figure 3.1 illustrates some critical events in the notification system. The most interesting event is when the patient is infected. However, it is very difficult, often impossible, to establish the exact time this event occurs, especially if the disease has a very long incubation period. The event closest to infection is the onset of disease, i.e. the first time when the patient has symptoms of the disease. The time of this event can also be difficult to reconstruct and if the physician reporting the case is unable to estimate the most probable time of onset, this information will be missing in the report. The only reliable time in this procedure is the date of registration, i.e. the date when the report arrives to SMI.

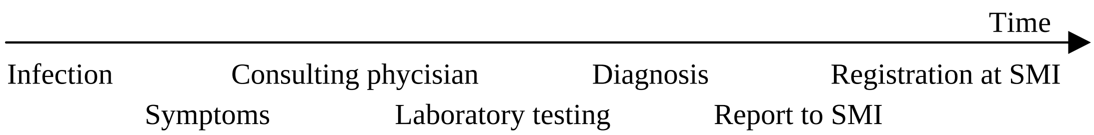

Figure 3.1. Critical events in the reporting chain.

Of course, there is a delay in reporting cases. In a register study of 20 selected notifiable infectious diseases, the time from disease onset to registration at SMI was examined (Jormanainen et al (1997)). Reporting delay was defined as the number of days between disease onset and registration at SMI. The median delay varied between diseases, from 15 days (meningococcal infection) to 91 days (atypical mycobacterioses) and was shorter for diseases that are more acute and longer for diseases of more chronic type. For campylobacter infections, the median delay was 19 days and within 64 days 95 % of the reports were registered. The distribution of the delay is shown in figure 3.2. A report with a delay of more than one year or less than one day are considered as miscoded and therefore excluded from further analysis. For this reasons 0.4 % of the reports are excluded from the analysis.

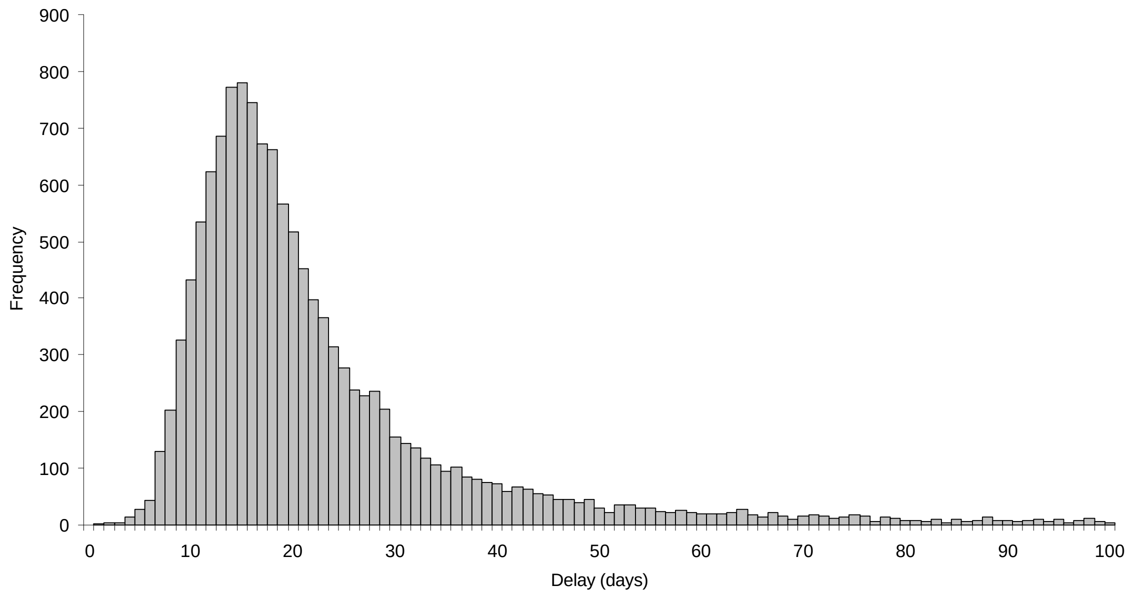

Figure 3.2. Reporting delay, i.e. the number of days between disease onset and registration of report at SMI for indigenous campylobacter infections 1992-1998.

Since we are studying variations of campylobacter infections in time we have to relate every case with a date. Because the disease onset date is missing from some of the reports, we are loosing observations if we want to use that as our time variable. About 89 % of the cases have information about date of onset. However, because of the relatively long delays and their skewed distribution, we will loose precision if we use the registration date. There are methods, e.g. back calculation, to estimate the missing onset dates from the registration dates. However, we will, in the following analysis, use the date of onset, without trying to recreate the missing observations.

3.3 Age and sex

Of the 13 077 domestic cases with information about date of onset between 1992-1998, about 53 % are men and 47 % women. The age distribution is the same for men and women with a high incidence among the youngest children and the young adults (20 - 35 years) see figure 3.3. For adults older than 35 years the incidence is decreasing with age. Notable is the dip in the incidence curve for children and youths between 5 and 20 years.

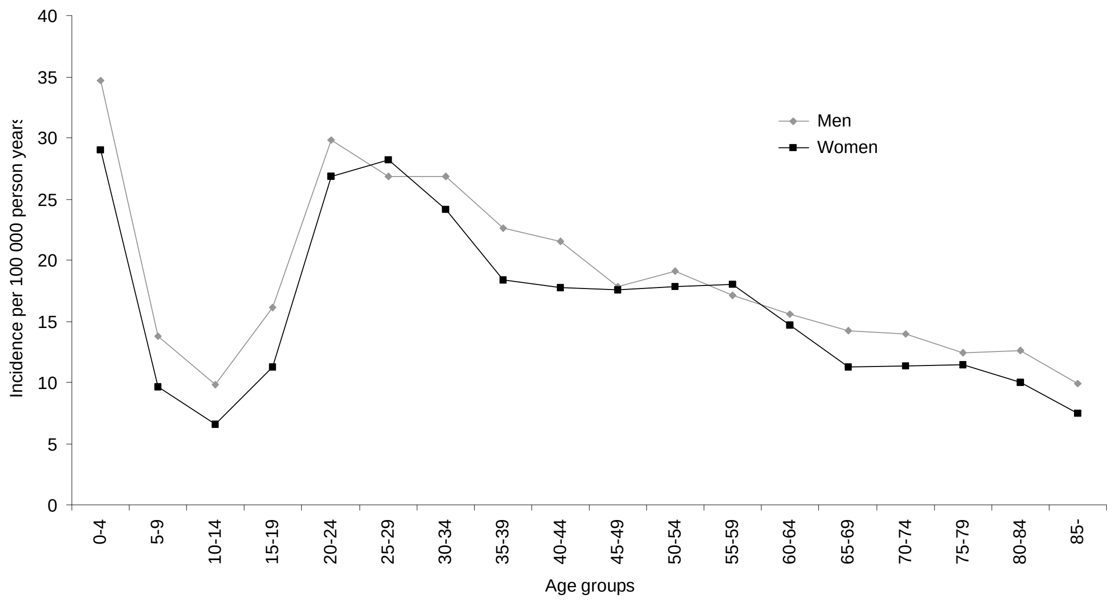

Figure 3.3. Age and sex distribution of incidence per 100 000 person-years of reported indigenous campylobacter infections, 1992-1998.

3.4 Geography

The geographical unit of the analysis are counties. These are the major administrative units of Sweden. The country is divided into 21 counties. The counties organise the health care within their area. In each of these counties, a county medical officer (smittskyddsläkare) is responsible for the local supervision of infectious diseases. In table 3.2 the number of indigenous cases and incidences per 100 000 inhabitants and year during the period under study are given for each county. The number of inhabitants is calculated as the mean of the number of inhabitants the last of December each of the years 1992 through 1998.

The incidence varies between 10 and 40 cases per 100 000 person-years. The extremes are Gotland with an incidence of 39.4 and Värmland with an incidence of 10.4. For most counties the population size is quite stable between and within years. However, for Gotland this is not the case. Gotland has the smallest population size. There are only about 58 000 inhabitants registered. Gotland is also one of the most popular counties to visit as a tourist and many people living in other counties in Sweden have their summerhouse there. This means that the actual population in Gotland is much higher in summer than in the rest of the year. The average number of visitors each year is approximately 600 000. This is a contributing cause to why Gotland has so much higher incidence than other counties. Another cause may be that the geological conditions of Gotland differ from the rest of the country in a way that influence the quality of the drinking water (cf. Andersson et al (1998)).

Table 3.2. Number of indigenous cases and incidence per 100 000 person-years of campylobacter infection in Sweden, date of onset 1992-1998. (For eight of the cases the sex was unknown.)

	County	Mean number of inhabitants (* 10^3)	Number of cases			Incidence per 100 000 person-years		
			Women	Men	Total	Women	Men	Total
1	**Stockholm**	1 717	1 171	1 351	2 523	16.6	20.2	18.4
2	**Uppsala**	285	249	226	475	21.5	20.1	20.8
3	**Södermanland**	258	190	216	406	18.3	21.1	19.7
4	**Östergötland**	413	232	261	493	14.0	15.9	14.9
5	**Jönköping**	328	206	240	447	15.6	18.4	17.0
6	**Kronoberg**	179	127	164	291	17.7	22.8	20.3
7	**Kalmar**	242	196	243	439	20.2	25.3	22.7
8	**Gotland**	58	84	98	182	36.0	42.8	39.4
9	**Blekinge**	152	85	86	171	14.0	14.2	14.1
10	**Skåne**	1 103	1 140	1 159	2 302	25.3	26.8	26.1
11	**Halland**	267	211	308	519	19.7	29.0	24.3
12	**Västra Götaland**	1 473	915	1 120	2 037	15.4	19.2	17.3
13	**Värmland**	283	106	130	236	9.3	11.6	10.4
14	**Örebro**	275	188	182	370	16.8	16.8	16.8
15	**Västmanland**	260	149	157	306	14.3	15.1	14.7
16	**Dalarna**	289	198	209	408	17.1	18.2	17.7
17	**Gävleborg**	287	126	144	270	10.9	12.6	11.7
18	**Västernorrland**	258	190	230	420	18.3	22.4	20.3
19	**Jämtland**	135	79	79	158	14.7	14.6	14.6
20	**Västerbotten**	258	119	154	273	11.5	15.0	13.2
21	**Norrbotten**	265	138	213	351	13.2	19.8	16.6
	Total	8 785	6 099	6 970	13 077	17.2	20.1	18.6

In figure 3.4 the counties and their incidences are indicated on a map of Sweden. At a first glance it is not easy to discern any geographical pattern. Maybe one can say that there is an over-representation of southern counties among those with higher incidence and an over-representation of northern counties among those with lower incidence.

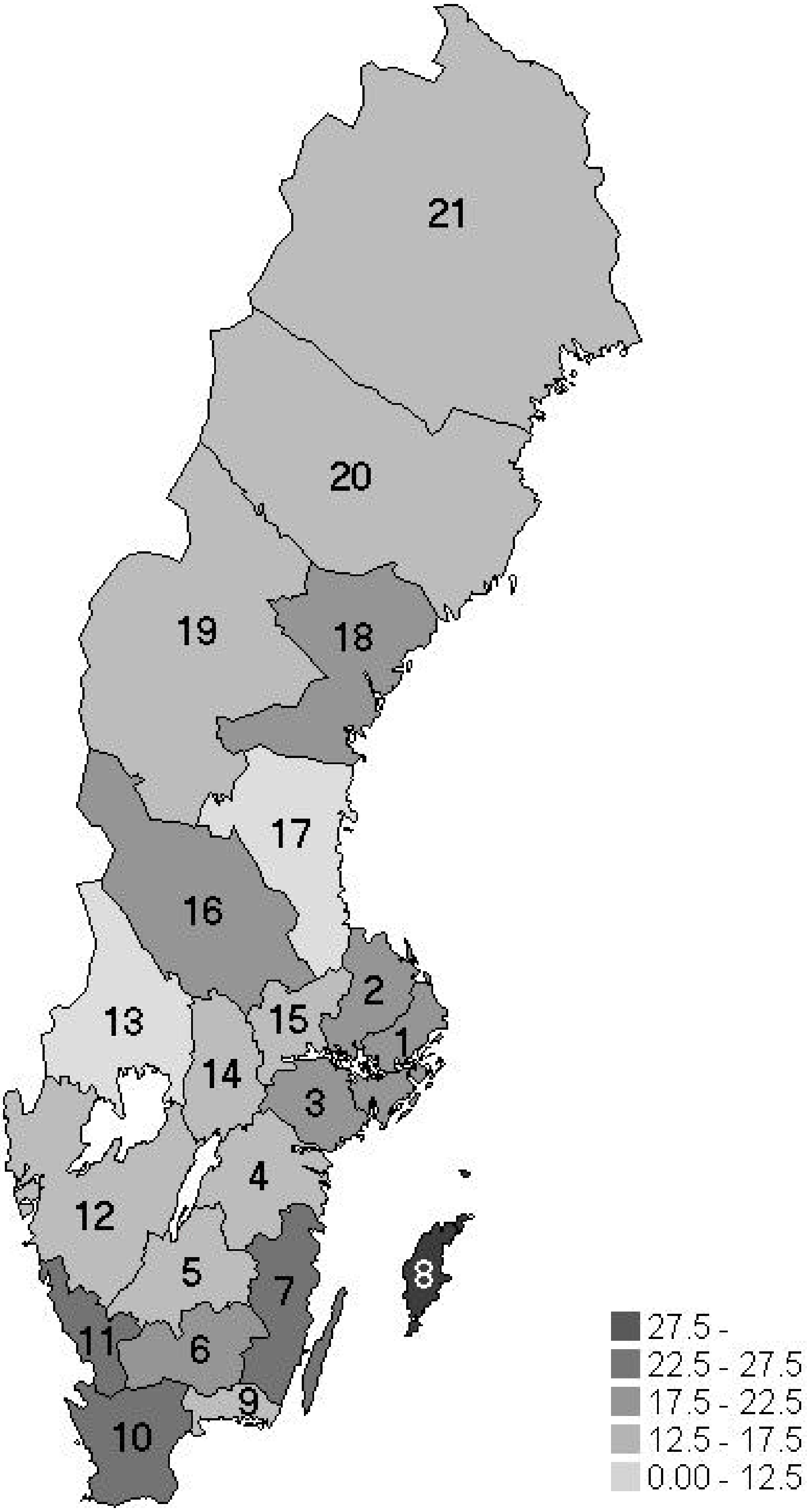

Figure 3.4. Map of Sweden indicating incidence per 100 000 person-years for each county.

3.5 Seasonability

In figure 3.5 the time series of the number of reported campylobacter infections (in the entire country) week by week during 1992-1999 is given. The series reveals a large variation in the number of cases during the year and a rather stable yearly pattern. There is a high incidence period peaking approximately at the end of July or beginning of August each year. The exact time of the peak and the duration of the high incidence period seem to vary between the years.

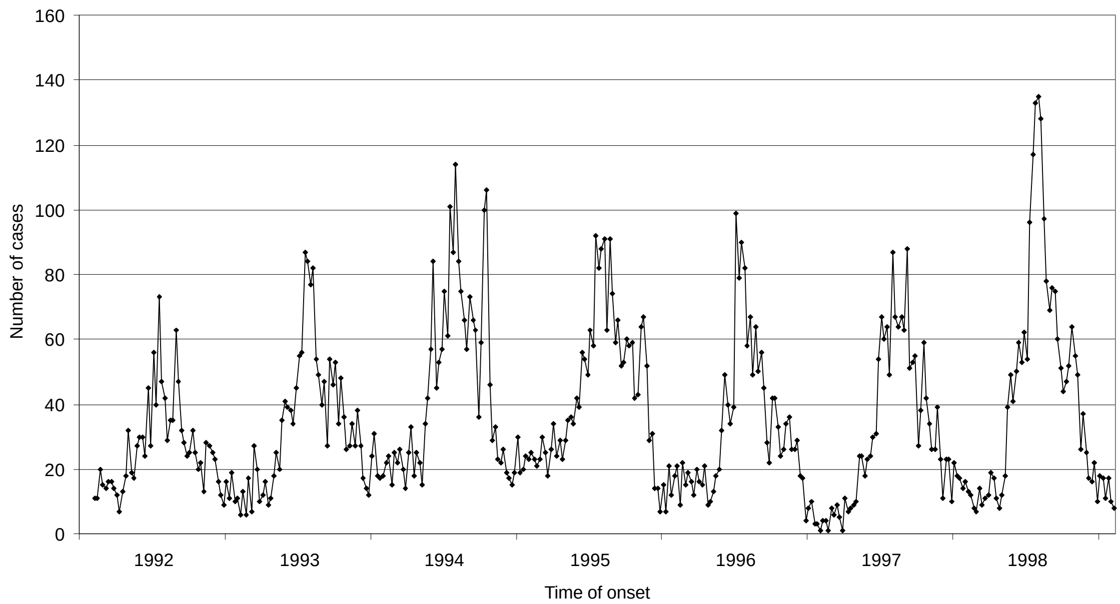

Figure 3.5. Weakly number of reported indigenous cases in Sweden. (Week no 6 1992 to week 5 1999).

This pattern is not only seen in the aggregated data but also in the time series for the separate counties. However, the time of the peak and the duration of the high incidence period vary between the counties. This is illustrated in figures 5.2-5.4 below. There the weekly number of reported cases is given for Stockholm, Blekinge and Jämtland together with a smooth estimate of the mean number of reports.

4 Analysis of temporal patterns

4.1 The data

The data used in the following analysis are the weekly number of reported cases by date of onset in each of the 21 counties. The time series start with week 6 in 1992 and end with week 5 in 1999. We have chosen to start the analysis with week 6 mainly for technical reasons. The smoothing model used defines a yearly parameter for the lowest incidence. It is thus convenient to relate a change of parameters to the time when the incidence is lowest. This happens around week 6. The time span analysed consists of 365 weeks. Accordingly, there are 365 observations for each county.

Reports with no date of onset of the disease have been excluded. The proportion of such reports is 11 %.

Data from the known large outbreaks (cf. section 2.3) have been modified to reflect a situation with only sporadic cases, i.e. the actual reported number of cases has been substituted by a mean number of reported cases in the weeks before and after the outbreak.

4.2 A crude analysis of temporal and spatial distribution of cases

The time series of cases (cf. Figure 3.5) shows large random variations. To be able to discern differences and to compare the patterns in different counties it is necessary to calculate statistics that illustrate the important features of the regular patterns. We will start by presenting some crude calculations, which are not based on any assumptions on the nature of the random variations or of the form of the temporal and spatial variations. The purpose of the analysis is to compare the time when the incidence peaks and when the high incidence period starts for the different counties.

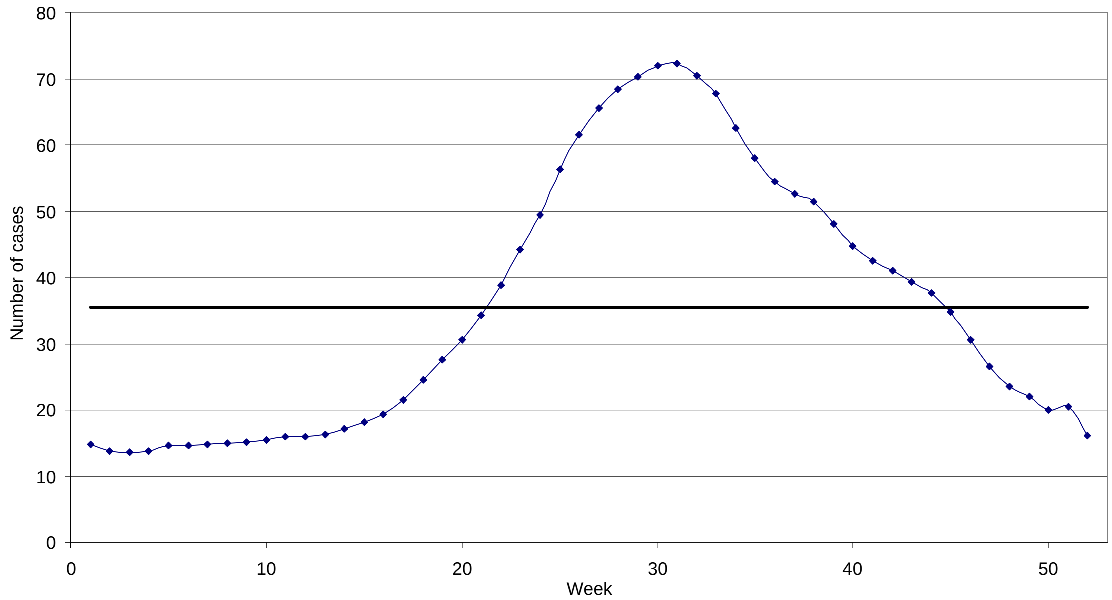

Figure 4.1. Nine-week centred moving average for the number of reported campylobacter cases averaged over years for Sweden. The straight line indicates the mean number of reported cases per week.

For each county the data has been aggregated on a yearly level. That is the number of cases occurring during i'th week within the year in the seven years under investigation have been summed and a mean number of cases per week have been calculated.. The reports from different years are aggregated to weeks with the same position within the year. Two statistics have been calculated. The first is the number of the mid-week in the 9-week period (of consecutive weeks) with the largest number of cases. The second statistic is the first week within the year of a consecutive period of 9 weeks which has an incidence larger that the yearly incidence. The calculations are illustrated in figure 4.1 where the 9-week moving average is given for data from Sweden.. The curve has its peek in week 31 and it crosses the line indicating the mean number of reports per week for the average number of reports during week 18 to 26.

Table 4.1 shows the result of the calculations for the individual counties.

Table 4.1. Peak week and start of high incidence period in the counties. The last column shows the relative rank of the north – south position of the residential cities

	County	Peak week	Start of high incidence period	Rank of N-S position
1	Stockholm	31	19	10
2	Uppsala	34	18	7
3	Södermanland	27	17	12
4	Östergötland	28	18	13
5	Jönköping	30	18	14
6	Kronoberg	31	18	17
7	Kalmar	30	18	18
8	Gotland	29	18	16
9	Blekinge	29	16	20
10	Skåne	31	17	21
11	Halland	32	19	19
12	Västra Götaland	31	18	15
13	Värmland	33	20	9
14	Örebro	30	17	11
15	Västmanland	28	18	8
16	Dalarna	32	18	6
17	Gävleborg	31	20	5
18	Västernorrland	32	20	4
19	Jämtland	32	21	3
20	Västerbotten	31	20	2
21	Norrbotten	32	19	1
	Sweden	31	18	

A general impression is that the peak week occurs earlier in the southern part of Sweden than in the northern part. The same seems to hold for the start of the high incidence period. This impression is to some extent confirmed by calculations of rank correlations between these numbers and the north-south position of the counties. The counties have been ordered according to the relative positions of their residential cities (cf. table 4.1). The rank correlation between peek week and the relative north-south position of the county is -0.42. The rank correlation for the start of the high incidence period is -0.65. The second of these rank correlations differs significantly from zero on the 5 % level, but not the first.

In the analysis made below with a more sophisticated parametric smoothing model a similar pattern occurs but the correlation with the north-south position does not differ significantly from zero.

4.3 Smoothing by MCMC-methods

There is evidently a large amount of randomness associated with the number of reported cases. In order to discern patterns as regards to variations within a year and between years we will have to smooth the time series in a convenient way. We will here do this by applying a model that describes how the mean number of cases varies in time and describes the random variations around this mean. The model that is used is in many respects very crude. It should not be regarded as a realistic stochastic model but rather as a model that produces a smoothed version of the time series. By studying the parameters in this smoothed version, we may get a better view of underlying regularities in the spread of campylobacter.

The smoothing will be done for the aggregated data from Sweden and for all separate counties. No effort has been made to make a simultaneous smoothing of the 21 counties. It is possible to do this by extending the model used, to consider spatial aspects. Such a model is sketched later and result from such an analysis will be presented later.

4.3.1 A model based on Poisson variation

The number of cases that occur during a particular week is assumed to be Poisson distributed with a mean that may depend on the week. The random variations in the different weeks are assumed to be stochastically independent. For several reasons it may be more appropriate to use a compound Poisson distribution to describe the random variation instead of a Poisson distribution. This is discussed in section 4.4.

The mean of the Poisson distribution will vary in time. In the following analysis, we will use a model that tries to capture some of the features that are seen in the crude time series by assuming a special parametric description of this variation. A model is formulated that accounts for

- that there is a flow of cases during the entire year,
- that within each year there exists one high incidence period, and that the time of the peak may vary between the years,
- that the duration of the high incidence period may vary between years,
- that the ratio between the incidence in the high and low incidence periods may vary between years.

The weeks are numbered successively from week no 5 (in February) 1992 onwards. This means that there are 365 observations (corresponding to weeks). The number of reported cases is denoted by $Y_1, Y_2 \ldots Y_{365}$.

According to assumption, these numbers are stochastically independent and Poisson distributed, i.e.,

$$Y_j \sim Po(m_j)$$

We have used the following model for the means:

$$\ln(m_j) = \ln(I) + v_j b_{yr} + (1 - v_j) b_{yr+1} + (T + t_{yr}) \left(\frac{\cos(2p(v_j - (\Theta + q_{yr})) + 1)}{2} \right)^{K + k_{yr}}$$

I is the mean number of reported cases per week during the entire period 1992-1999. The term ln(I) serves as normation. The purpose of it is to make the parameters comparable between the counties and is otherwise redundant in the model.

The index yr stands for the year to which week j belongs and v_j stands for the relative position of the week within that year:

The parameters b_{yr} describe the incidence of sporadic cases during the low incidence period of the year.

In fact

$$\mathrm{I}\exp(\mathrm{v}_j \mathrm{b}_{\mathrm{yr}} + (1 - \mathrm{v}_j)\mathrm{b}_{\mathrm{yr+1}})$$

is the incidence of reported cases during a low incidence period in year yr. In order to make this expression smooth we have used this polygonal function that changes continuously over the year instead of a jump function that changes value for each year.

The expression

$$(T + t_{yr}) \left(\frac{\cos(2p(v - (\Theta + q_{yr})) + 1}{2} \right)^{K + k_{yr}}$$

describes the variation of the incidence over a year. It takes its highest value in year yr at time $2p(\Theta + q_{yr})$. At that time, the incidence is $\exp(T + t_{yr})$ higher than during a low incidence period during that year.

The duration of the high incidence period is related to $K + k_{yr}$. A high value gives a short duration of the high incidence period and a low value a long high incidence period.

The model is formulated in such way that parameters T, K and Θ are general parameters that are uninfluenced by the year. The indexed parameters τ_{yr}, κ_{yr} and θ_{yr} measure the yearly deviance from the values of the general parameter.

The statistical analysis is performed within a Bayesian framework. This means that we will apply prior distributions to the parameters and use the means of the posterior distributions as estimates of the parameters. The priors used will all be vague. This guarantees that the posterior distributions depend essentially on the observed data. The choice of the prior distributions are standard (cf. Gilks et al (1996)).

The following prior distributions have been used:

$b \propto N(0,100),$
$T \propto Gamma(0.1,0.1),$
$Q \propto U(0.5,2p-0.5),$
$K \propto Gamma(0.08,0.08),$
$t_{yr} \propto N(0,s_t),$
$q_{yr} \propto N(0,s_q),$
$k_{yr} \propto N(0,s_k),$
$s_q \propto U(0.00001,p/2),$
$s_t \propto U(0.00001,2),$
$s_k \propto U(0.00001,2).$

The structure of the model is illustrated in the form of a “doodle” in figure 4.2.

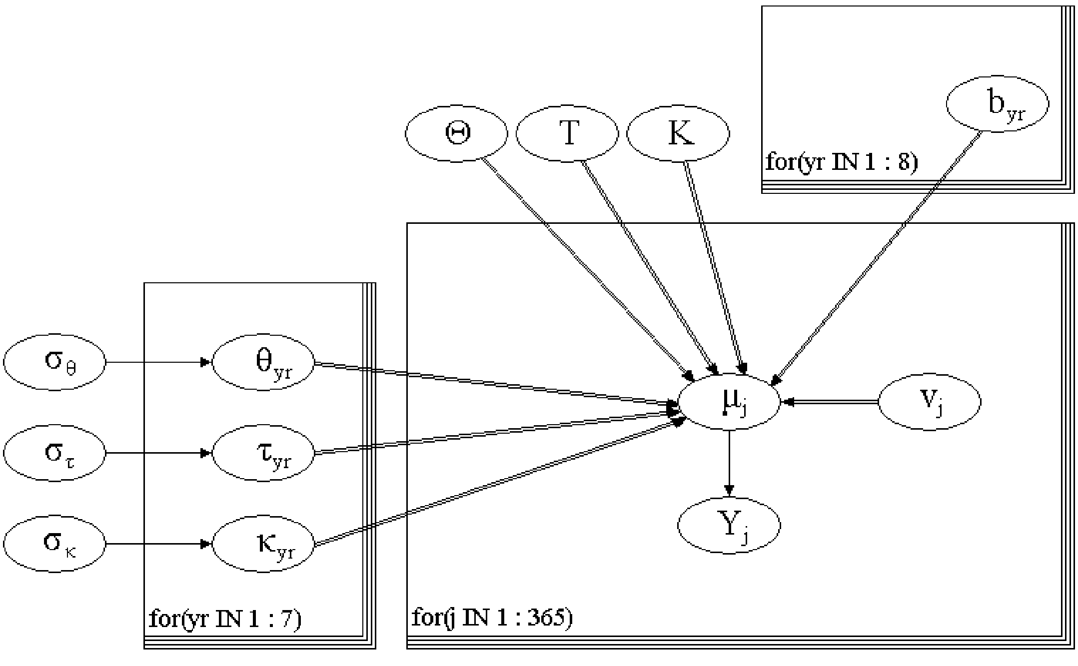

Figure 4.2. Sketch of the model.

4.3.2 Estimation of parameters

The parameters in the model are estimated with Bayes-estimates. This means that the posterior distribution for the parameters given the observed data and assumed prior distributions are calculated. The calculation of posterior distributions is a difficult and heavy computational task. However, using modern simulation techniques this can be done with sufficient accuracy and speed. The calculations are carried through with help of an MCMC-algorithm (cf. Gilks et al (1996)) provided by the programme BUGS (cf. Spiegelhalter et al (1999)). The simulations are done with the Metropolis-Hastings algorithm.

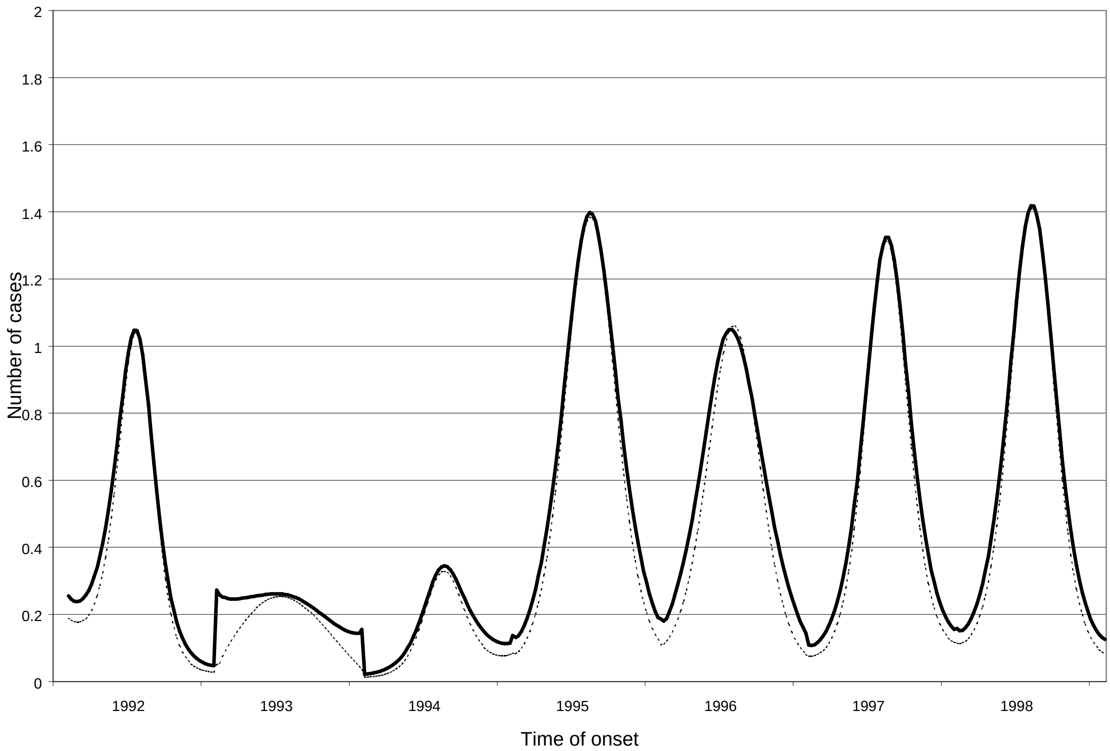

Figure 4.3. Illustration of the two methods for estimating weekly incidence based on data from Blekinge. Dashed line represents the model derived from the estimated basic parameters. Bold line represents the estimated parameter μ.

The estimates of the parameters are the means of the corresponding (simulated) posterior distributions. However, we are not only interested in the values of the individual parameters but also in the function describing the variation of the incidence in time. The incidence for a particular week, j, can be estimated in several ways. One possibility is to estimate the basic parameters b_{yr}, T, τ_{yr}, K, κ_{yr}, Θ, θ_{yr} and use these estimates in the formula for $\ln(\mu_j)$ which defines the model. Another possibility is to estimate the incidence for each week (μ_j) with the mean of the posterior distribution. In general, these two estimates will be different. In particular, this will be the result if the posterior distributions of the parameters have heavy tails. In the data we have analysed, the difference between the two ways of estimating the variation of the incidence in time will be very small. In some cases where the analysis is based on very few reported cases there is a substantial difference. This is exemplified in figure 4.3 where the two methods have been applied to the data from Blekinge. Observe the difference that appears for 1993.

In the figures presented below, we have chosen to use the first method for estimating the smoothed incidences for the weeks. In this way the estimated curves that are presented will have the functional form implied by the formula for $\ln(\mu_j)$.

4.4 The Poisson assumption and overdispersion

In the model discussed in the previous section, we have assumed that the number of reported cases each week is Poisson distributed. This assumption can be questioned for several reasons. One important feature of the Poisson distribution is that its mean equals its variance, i.e., the ratio between the variance and the mean is 1. A distribution with a ratio greater than 1 is called overdispersed. There are strong reasons to believe that an overdispersed distribution should be more appropriate than the Poisson distribution. Overdispersed distributions can be motivated by the presence of

- Minor outbreaks

The campylobacter cases may occur simultaneous in several persons due to exposure to the same infectious source. Even if we have discarded reported cases from major outbreaks from our analysis, there still may be reported cases from minor outbreaks. This should imply that there is dependence between cases. The most natural assumption is that the number of events when campylobacter infections are transmitted is Poisson distributed and the number of persons infected in such event is a random. The number of reported cases should in that case be modelled as a sum of a Poisson distributed number of independent random number of cases. This will yield an overdispersed distribution.

- Secondary cases

An infected person may spread the infection further to individuals in the neighbourhood. It is often claimed in the literature that such secondary infections are uncommon. A mechanism of this kind would also cause some cases to depend on each other and imply an overdispersed distribution rather a Poisson distribution. Another consequence could be a (stochastic) dependence of observations in subsequent weeks.

- Dependent reporting

Dependencies between reported cases can also be the effect of the reporting system. If the reporting of cases is not done independently but e.g., the reporting medical officer reports several unrelated cases simultaneously an artificial clustering, as relates reporting date, of cases occurs. Since we have chosen to use the date of the onset of the disease as the time associated with the infection we have possibly avoided this effect.

- Heterogeneity

Another possible cause of overdispersion is that the relatively large areas for which the data are presented in fact consist of several sub-areas that have different patterns as regards to the temporal variation in the number of campylobacter cases. Such sub areas could, e.g. be rural and urban areas or coastal and inland parts of a county.

- Thinning due to underreporting

It is well established that the campylobacter cases are severely underreported. A simple model for underreporting is that each case is reported with a certain probability, independent of other cases being reported. The observed series of reported cases is then a thinned version of the series of all campylobacter infections. A theoretical analysis shows that this kind of underreporting results in observations that are more Poisson-like than the unthinned series. Thus underreporting has an opposite effect compared to the other problems mentioned since it will make the distribution less overdispersed.

Table 4.2. Estimated overdispersion

	County	Overdispersion
1	Stockholm	1.79
2	Uppsala	1.12
3	Södermanland	1.17
4	Östergötland	1.19
5	Jönköping	1.35
6	Kronoberg	0.96
7	Kalmar	1.27
8	Gotland	1.08
9	Blekinge	1.29
10	Skåne	1.58
11	Halland	1.36
12	Västra Götaland	1.48
13	Värmland	1.03
14	Örebro	1.05
15	Västmanland	1.17
16	Dalarna	1.13
17	Gävleborg	1.50
18	Västernorrland	1.37
19	Jämtland	1.25
20	Västerbotten	1.47
21	Norrbotten	1.09
	Sweden	2.99

In order to get an idea of the size of a possible overdispersion we can calculate the Pearson χ^2-statistic, which is a measure of how well the model fits the data.

$$P = \sum_i \frac{(Y_i - \hat{m}_i)^2}{\hat{m}_i}$$

If the random variation really is Poisson distributed and if the parameters are sufficiently well estimated this statistics should approximately equal the number of observations. In table 4.2 P/365 are given for the individual counties and for the aggregated data from Sweden. Due to heterogeneity the overdispersion for Sweden is, as expected, larger than for the individual counties.

5 Results

5.1 The fit of the model

The model described above has been applied both to the aggregated data from Sweden and to the individual counties. Estimates of all parameters are shown in the appendix (cf. tables A.1 to A.5). In figure 5.1, the estimated mean number of reported cases due to the model is given together with the actual number of cases reported for all of Sweden. It is seen that the model fits the data well.

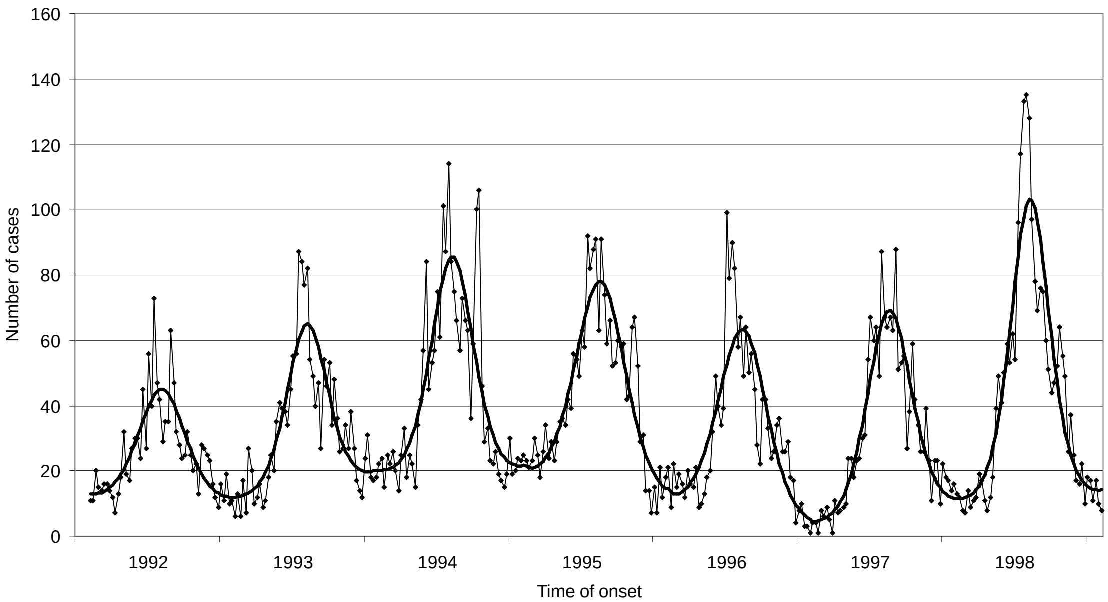

Figure 5.1. Observed and estimated mean number of cases week by week for Sweden.

It turns out that the model fits the data well also when data from individual counties are analysed. As examples, the results of the model applied to Stockholm (figure 5.2), Blekinge (figure 5.3) and Jämtland (figure 5.4) are shown below.

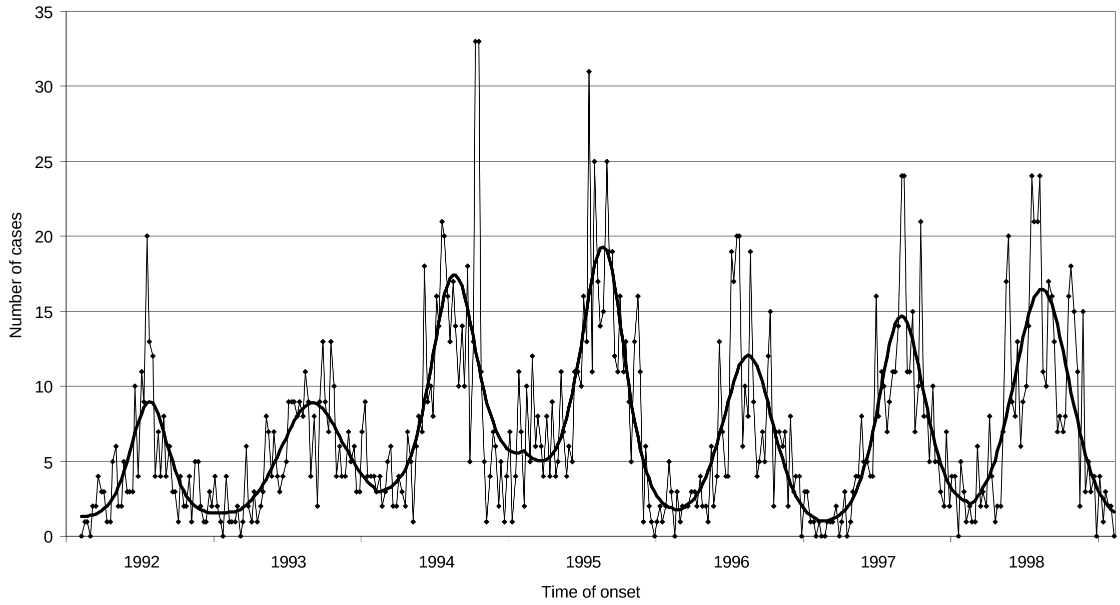

Figure 5.2. Observed and estimated mean number of cases week by week for Stockholm.

Since there are fewer cases in Stockholm than in Sweden the random variation around the estimated mean curve is larger in figure 5.2 than in figure 5.1. It should be observed that even if the pattern is rather similar in Stockholm and in the aggregated data from Sweden there still are some differences that are revealed in the observations and in the fitted model.

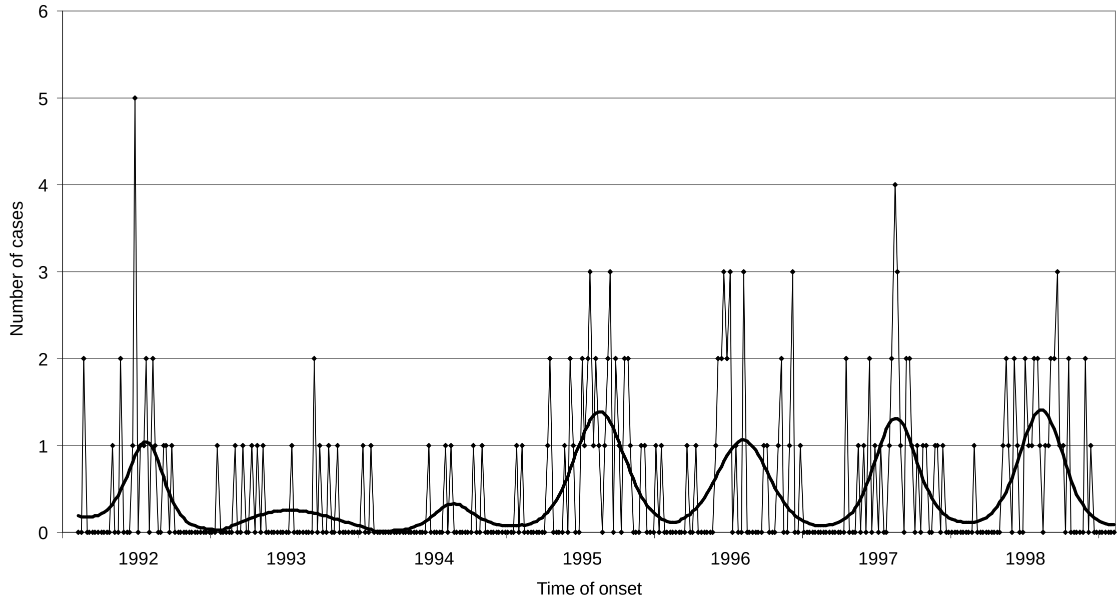

Figure 5.3. Observed and estimated mean number of cases week by week for Blekinge.

The number of cases in Blekinge is small. During the entire period consisting of 365 weeks there were only 171 cases reported. In most of the weeks there were 0 or 1 case reported. It may seem surprising that the model despite of this is able to discern a regular pattern. However, observe that the peaks in 1993 and 1994 for Blekinge are not very prominent.

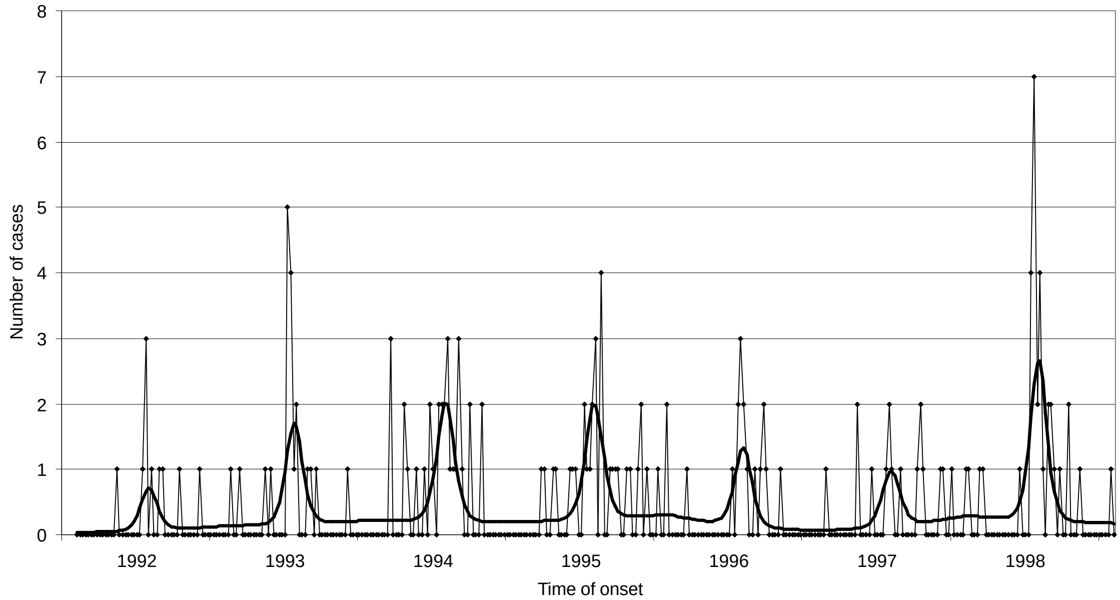

Figure 5.4. Observed and estimated mean number of cases week by week for Jämtland.

Despite of the number of reported cases also in Jämtland is small the model succeeds in finding a regular pattern in the observed time series with very marked peaks.

5.2 Some results

5.2.1 National

A closer look at the parameter values obtained in the analysis for the entire country indicates that peak incidence occurs in week 32. The peak varies between years from week 31 until week 33 (cf. Table 5.1).

Table 5.1. Estimated peak week according to the model

Year	1992	1993	1994	1995	1996	1997	1998
Peak week	31	31	32	33	33	31	31

The amplitudes (T, and $T+\tau_{yr}$) measure the logarithm of the ratio of the highest incidence in the model to the underlying basic incidence. The yearly variation of the ratio between high and low incidences is given in table 5.2.

Table 5.2. The ratio between high and low incidences

Year	1992	1993	1994	1995	1996	1997	1998
Ratio	3.6	4.2	4.1	4.6	8.4	9.4	8.1

In order to illustrate when the high incidence period starts we have calculated the number of the week in which the incidence is twice the basic low incidence. A simple calculation yields that this will be week no

$$\frac{365}{7}\left((\Theta+\theta_{yr})-\arccos\left(2\left(\frac{\ln(2)}{(T+\tau_{yr})}\right)^{1/(K+\kappa_{yr})}-1\right)\right)$$

The estimated values are given in table 5.3.

Table 5.3. Week when the 2-times-base incidence period starts

Year	1992	1993	1994	1995	1996	1997	1998
Starting week	20	20	20	19	16	15	17

5.2.2 By county

When the analysis is made for each county, it is confirmed that there are substantial differences between the 21 counties.

For each county we have calculated estimates of the peak week, the ratio between the highest and lowest incidence and the start of the 2-times-base incidence period (cf. table 5.4). Calculations of the rank correlations between these estimates and the north-south position of the county (cf. table 4.1) only yield a correlation that differs significantly (on the 5 % level) from zero for the ratio between highest and lowest incidence. The ratio seems to be larger in the northern part of the county than in the southern. However, it should be observed that some counties tend to have a ratio that differs remarkably from its neighbours (e.g. Blekinge and Sörmland).

Tables A.6 to A.8 in the appendix, give the estimated values of the peak week, the ratio of high and low incidences and the start week of the 2-times-base incidence period for the 21 counties and the 7 years:

Table 5.4. Peak week, ratio between high and low incidence and start of 2-times-base incidence period for the 21 counties

	County	Peak week	Ratio between high and low incidence	Start of 2 times base incidence period
1	**Stockholm**	33	6.7	16
2	**Uppsala**	33	4.6	17
3	**Södermanland**	31	9.6	14
4	**Östergötland**	33	5.3	16
5	**Jönköping**	31	4.6	20
6	**Kronoberg**	33	7.4	16
7	**Kalmar**	29	3.4	21
8	**Gotland**	30	5.5	21
9	**Blekinge**	32	12.9	10
10	**Skåne**	32	4.4	19
11	**Halland**	34	4.6	20
12	**Västra Götaland**	33	5.4	17
13	**Värmland**	32	5.7	21
14	**Örebro**	31	6.6	15
15	**Västmanland**	30	7.6	19
16	**Dalarna**	32	7.1	17
17	**Gävleborg**	31	9.2	18
18	**Västernorrland**	31	8.9	23
19	**Jämtland**	31	9.3	24
20	**Västerbotten**	32	8.4	21
21	**Norrbotten**	33	11.8	18

6 Conclusions

The quality of the reporting of campylobacter infections is obviously low. A large proportion of cases are never reported. This is true both for cases that occur as part of an outbreak and those that are sporadic events. In spite of this, a study of the statistics may tell us something of the spread of campylobacter infections. The investigation made above verifies that the number of reported (sporadic) cases shows a very stable pattern, both as regards their distribution in time and in space. It is not probable that this regularity is an artefact of a poor reporting system. Every summer there is a high incidence period, which starts, with some variation, at approximately the same week each year. The highest incidence each year occurs, also with some variation, at approximately the same time every year. The developments of the number of reported infections in the different counties are qualitatively similar in the twenty-one counties of Sweden. However, there are some geographical variations in placement, duration and strength of the high incidence period.

The purpose of this paper is to illustrate the regularities and to illuminate the variations between statistics from different counties in order to make comparisons and analysis possible. Eventually this may lead to a better understanding of the mechanisms for transmission of campylobacter infections. A main tool is a model for smoothing of the raw statistics of reported cases. The model seems to fit the observed data very well, even if there are some signs of overdispersion. This makes it possible to compare statistics from different counties by comparing the corresponding estimate of the parameters (or functions of them) of the model. The study presented is a starting point for further analysis that associates the variations of the parameters to different covariates. In order to do that, the statistical models have to be developed further.

7 Plans for future research

The model used in the analysis has to be viewed as a first attempt that catches some important features of the raw time series of reported campylobacter infections. There are other features, which should be incorporated in a more advanced model. It is also of interest to develop models that make it possible to study the influence of covariates on variations between counties and years. A few examples of such covariates are geography (e.g. north-south position of the county), weather (e.g. precipitation), and auxiliary measurements of campylobacter incidence in animal populations (e.g. campylobacter prevalence in chicken).

Once the basic model has been established such developments are fairly straightforward. A simultaneous analysis of data from all counties should retain the structure of the model for each county but also allow for smoothing between geographically close countries. The influence of covariates can be accounted for by using models for how these influences parameters values. Future research involves formulation of models that make such analysis possible and the development of technical aids to do the necessary calculations.

Some crude calculations implied that there is more than Poisson variation in the data. The model can be developed to include a “scale” parameter describing this overdispersion, by assuming negative binomial variation instead of Poisson variation.

This paper only deals with campylobacter infections. The approach used could of course be applied to study the spatial and temporal variations of the number of reports of other infections. Each disease has a pattern of its own. Thus, it is possible that the model has to be modified depending on the particular circumstances that are important for the spread of each particular infection. One important feature for many diseases of interest that we have disregarded in the analysis of campylobacter infections is man-to-man spread. The reason is that it is the common understanding that such transmission only plays a small part. For other infections, such spread can be essential. Transmission of infections between individuals will create dependencies between reported cases both in time and in space. It is thus of interest to develop models that take dependencies into account.

Acknowledgement

We wish to thank Yvonne Andersson, Birgitta de Jong and Prof. Johan Giesecke for sharing their great knowledge about the epidemiology of campylobacter infections and for giving valuable comments on this report.

This work was funded by The Bank of Sweden Tercentenary Foundation.

8 References

Andersson, Y., Bast, S., Gustavsson, O., Jonsson, S., & Nilsson, T. (1994). Campylobacterutbrott [A Campylobacter outbreak]. *Epid-aktuellt, 17*, 6-9. (In Swedish)

Andersson, Y., & Gustavsson, O. (1998). *Campylobacter – en vattenburen smitta att beakta vid planläggning av vattentäkt av ytvattentyp [Campylobacter – an important source of waterborne infections to pay attention to projecting a surface water supply]* (Forskningsrapport). Stockholm: Överstyrelsen för civil beredskap. (In Swedish)

Andersson, Y., Gustavsson, P., Hammarquist, C., Kulander, L., Nilsson, P. E. W., & Olsson, A.-M. (1998). *Vattenkvalitet och hälsa, en studie av självrapportering vid mag- och tarmsjukdom [Drinking water quality and health. A study on self-reporting of gastrointestinal illness]*. Visby: Godrings tryckeri AB. (In Swedish)

Andersson, Y., Normann, B., & Tideström, L. (1999). *Fakta om smittsamma sjukdomar från A till Å*. Stockholm: Smittskyddsinstitutet. (In Swedish).

Bresky, B., Studahl, A. C., Säll, C., & Luther, M. L. (1995). Stort utbrott av magsjuka i Marks kommun [A large outbreak of gastrointestinal illness in the municipality of Mark]. *Epid-aktuellt, 18*, 6-8. (In Swedish)

Chin, J. (Ed.). (2000). *Control of communicable diseases manual* (17^{th} ed.). Washington, DC: American Public Health Association.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Jormanainen, V., Lindbäck, J. & Giesecke, J. (1997). Hur länge är kliniska anmälningar på väg till Smittskyddsinstitutet? [Delay of clinical notifications]. *Smittskydd, 4*, 34-35. (In Swedish)

Kapperud, G. (1995). *Descriptive and analytic epidemiology of Campylobacter enteritis in humans. WHO consulting on Emerging Foodborne Diseases, Berlin, Germany, 20-24 March 1995*. World Health Organization. Federal Institute for health Protection of Consumers and Veterinary medicine.

Lighton, L. L., Kaczmarski, E. B. & Jones, D. M. (1991). A study of risk factors for Campylobacter infections in late spring. *Publ Hlth, 105*, 199-203.

Lindbäck, J. *Geografisk spridning av campylobacterinfektioner, statistisk analys med MCMC [Geographical spread of campylobacter infections, statistical analysis using MCMC]* (Forskningsrapport. Serie B. Tillämpad forskning., nr B:52). Stockholm: Stockholms universitet, Institutet för försäkringsmatematik och matematisk statistik. (In Swedish)

Lindqvist, R., Andersson, Y., Lindbäck, J., Wegscheider, M., Tideström, L., & Eriksson, Y. (1999). *MATUPP – intesivstudie av matförgiftningar i Uppsala kommun under ett år [An intensive study of food poisoning in the municipality of Uppsala]* (Livsmedelsverkets rapport., nr 12). Uppsala: Livsmedelsverket. (In Swedish)

Notermans, S. (1995). *Epidemiological aspects of thermophilic Campylobacter in water related environments. WHO consulting on Emerging Foodborne Diseases, Berlin, Germany, 20-24 March 1995.* World Health Organization. Federal Institute for health Protection of Consumers and Veterinary medicine.

Smittskyddsinstitutet. (1995). *Smittsamma sjukdomar i Sverige 1994, Epidemiologiska enhetens årsrapport. [Infectious diseases in Sweden 1994, yearly report of the dept. of epidemiology]* Stockholm: Smittskyddsinstitutet. (In Swedish)

Smittskyddsinstitutet. (1996). *Smittsamma sjukdomar i Sverige 1995, Epidemiologiska enhetens årsrapport. [Infectious diseases in Sweden 1995, yearly report of the dept. of epidemiology]* Stockholm: Smittskyddsinstitutet. (In Swedish)

Smittskyddsinstitutet. (1997). *Smittsamma sjukdomar i Sverige 1996, Epidemiologiska enhetens årsrapport. [Infectious diseases in Sweden 1996, yearly report of the dept. of epidemiology]* Stockholm: Smittskyddsinstitutet. (In Swedish)

Spiegelhater, D. J., Thomas, A. & Best, N. G. (2000). *WinBUGS Version 1.3 User Manual.* MRC Biostatistics Unit, Cambridge.

Tauxe, R. V. (1992). Epidemiology of Campylobacter jejuni infections in the United States and other industrialized nations. In: Nachamkin, I., Blaser, M.J., Tompkins, L.S. (Eds): *Campylobacter jejuni. Current status and future trends*. American Society for Microbiology, Washington DC.

Wheeler, J. G., Sethi, D., Cowden, J. M., Wall, P. G., Rodrigues, L. C., Tompkins, D. S., Hudson, M. J. & Roderick, P. J. (1999). Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *BMJ, 318,* 1046-1050.

Appendices

Table A.1. Estimates of T, K, Θ and 95 % Bayesian credibility intervals.

		T	**(95 % c.i.)**	**K**	**(95 % c.i.)**	**T**	**(95 % c.i.)**
1	**Stockholm**	1.90	(1.50 ; 2.32)	1.00	(0.51 ; 1.75)	3.35	(3.11 ; 3.56)
2	**Uppsala**	1.52	(0.024 ; 2.35)	0.88	(0.003 ; 2.10)	3.37	(2.98 ; 3.72)
3	**Södermanland**	2.26	(1.62 ; 3.20)	0.89	($1\cdot10^{-6}$; 1.82)	3.18	(2.80 ; 3.66)
4	**Östergötland**	1.67	(0.74 ; 2.81)	0.73	($3\cdot10^{-6}$; 1.77)	3.37	(3.00 ; 3.69)
5	**Jönköping**	1.52	(1.14 ; 1.95)	1.65	(0.71 ; 3.26)	3.18	(2.91 ; 3.45)
6	**Kronoberg**	2.00	(1.27 ; 2.86)	0.84	(0.23 ; 1.75)	3.39	(3.03 ; 3.68)
7	**Kalmar**	1.23	($1\cdot10^{-6}$; 1.93)	3.96	(1.58 ; 8.00)	2.88	(2.68 ; 3.28)
8	**Gotland**	1.70	(1.07 ; 2.26)	3.32	(1.23 ; 6.54)	2.99	(2.76 ; 3.23)
9	**Blekinge**	2.56	(1.45 ; 3.92)	0.50	($1\cdot10^{-17}$; 2.34)	3.25	(2.89 ; 3.60)
10	**Skåne**	1.49	(0.96 ; 2.00)	1.37	(0.95 ; 1.81)	3.24	(3.04 ; 3.46)
11	**Halland**	1.53	(1.14 ; 1.94)	1.24	(0.49 ; 2.25)	3.45	(3.21 ; 3.70)
12	**Västra Götaland**	1.68	(1.21 ; 2.20)	1.12	(0.77 ; 1.58)	3.35	(3.16 ; 3.60)
13	**Värmland**	1.74	(0.95 ; 2.40)	2.64	(1.19 ; 4.54)	3.28	(3.00 ; 3.71)
14	**Örebro**	1.89	(1.36 ; 2.63)	0.95	(0.30 ; 1.78)	3.18	(2.92 ; 3.45)
15	**Västmanland**	2.02	(1.51 ; 2.59)	2.76	(0.42 ; 5.78)	3.02	(2.81 ; 3.28)
16	**Dalarna**	1.96	(1.48 ; 2.48)	1.19	(0.02 ; 2.16)	3.27	(3.05 ; 3.53)
17	**Gävleborg**	2.22	(1.60 ; 2.92)	1.94	($1\cdot10^{-4}$; 4.01)	3.13	(2.91 ; 3.31)
18	**Västernorrland**	2.19	(1.77 ; 2.56)	5.00	(2.60 ; 7.97)	3.10	(2.93 ; 3.29)
19	**Jämtland**	2.23	(1.65 ; 2.80)	7.01	(2.00 ; 19.1)	3.13	(2.96 ; 3.35)
20	**Västerbotten**	2.13	(1.58 ; 2.70)	2.46	(0.93 ; 4.93)	3.29	(2.83 ; 3.77)
21	**Norrbotten**	2.47	(1.95 ; 3.03)	1.41	(0.009 ; 2.51)	3.33	(3.12 ; 3.53)
	Sweden	1.65	(1.35 ; 2.00)	1.17	(0.88 ; 1.44)	3.31	(3.19 ; 3.49)

Table A.2. Estimates of $T+\tau_{yr}$ and the standard deviation, σ_τ, of τ_{yr}.

		1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**s t**
1	**Stockholm**	1.8518	1.3902	1.4380	1.8663	2.2079	2.2472	2.2278	0.4984
2	**Uppsala**	1.7347	1.6198	1.8411	2.1634	2.0757	2.1545	2.2304	0.4436
3	**Södermanland**	1.9965	2.0206	1.7038	1.8297	1.9895	1.8741	2.0000	0.3229
4	**Östergötland**	1.2857	2.3175	2.0831	1.7595	2.4312	1.8482	1.9777	0.6214
5	**Jönköping**	1.7782	1.9453	1.7928	1.9038	1.9426	2.0239	1.9720	0.2169
6	**Kronoberg**	1.7289	1.9407	1.6551	1.6120	2.2581	2.0547	2.1475	0.4542
7	**Kalmar**	1.6227	1.5378	1.9132	2.1253	2.4719	2.2609	2.4806	0.6295
8	**Gotland**	1.7607	2.0971	1.7704	2.1705	1.8131	1.9631	1.9556	0.3852
9	**Blekinge**	1.9756	1.9523	1.6257	2.0233	1.8104	1.9932	2.0258	0.3991
10	**Skåne**	1.5627	1.7515	1.9692	1.3553	1.9628	2.5783	2.1411	0.5448
11	**Halland**	1.8198	1.7110	1.8920	1.8967	1.9548	2.0594	2.0502	0.2625
12	**Västra Götaland**	1.3587	1.7005	1.4752	1.6660	2.2588	2.4587	2.4592	0.6186
13	**Värmland**	2.1042	1.9150	1.9442	1.4298	1.7882	2.0027	2.4832	0.5680
14	**Örebro**	1.8489	1.9353	1.7850	1.7928	1.9797	1.8943	2.1189	0.2801
15	**Västmanland**	1.8034	1.8317	1.9244	1.7466	1.9188	2.0006	2.1896	0.3240
16	**Dalarna**	1.8602	1.8971	1.8880	1.7523	2.0149	1.8323	2.1576	0.3000
17	**Gävleborg**	2.0476	1.5085	1.6865	1.8216	2.2893	2.0209	2.0565	0.4797
18	**Västernorrland**	2.0028	1.7453	2.0444	1.7886	1.9345	1.8297	2.0383	0.2785
19	**Jämtland**	2.0166	1.9694	1.9369	1.7570	1.9298	1.6452	2.1309	0.3775
20	**Västerbotten**	1.7134	1.9336	1.8927	2.0706	1.7839	1.8228	2.1651	0.3563
21	**Norrbotten**	1.8799	2.0449	1.8195	1.7936	1.8406	1.8603	2.1633	0.3029
	Sweden	1.2894	-0.9189	-1.0191	-0.3665	0.2866	-0.0445	-1.3152	0.7285

Table A.3. Estimates of $K+\kappa_{yr}$ and the standard deviation, σ_{κ}, of κ_{yr}.

		1992	1993	1994	1995	1996	1997	1998	s ?
1	**Stockholm**	1.5860	0.7022	1.3475	1.2265	0.8027	0.9602	0.6454	0.5435
2	**Uppsala**	1.1133	1.2011	1.0384	0.7802	1.0978	1.2571	1.2224	0.5066
3	**Södermanland**	1.1135	1.1026	1.1010	0.8860	1.0063	1.5001	1.2680	0.4629
4	**Östergötland**	0.9594	1.0388	1.0485	0.9683	1.0862	0.9690	1.2406	0.3088
5	**Jönköping**	1.0733	1.1351	0.7849	0.9809	0.8835	1.0942	1.0774	0.4782
6	**Kronoberg**	0.9942	1.0327	1.0939	0.9086	1.1072	1.0664	1.0137	0.2867
7	**Kalmar**	0.4121	0.2349	1.0861	1.1301	1.3829	1.8877	1.2203	1.1670
8	**Gotland**	0.8896	0.2013	1.5065	1.1337	1.6769	0.9545	1.1673	1.0640
9	**Blekinge**	1.8625	0.6915	1.7499	1.1603	1.1741	1.4706	1.3713	0.8188
10	**Skåne**	0.9215	1.1866	1.0747	0.7304	1.0586	1.0335	1.0904	0.3181
11	**Halland**	0.7907	1.2204	1.0728	0.8800	1.0287	1.1084	1.0960	0.4243
12	**Västra Götaland**	1.0221	1.0128	1.0047	1.0461	0.8429	1.0180	1.0647	0.2417
13	**Värmland**	1.0346	0.9843	1.4658	0.9204	1.4181	0.5266	1.0148	0.8043
14	**Örebro**	0.8963	1.1121	0.9275	1.0513	1.0682	1.0630	1.0792	0.3380
15	**Västmanland**	1.2611	1.5382	1.1807	0.7290	0.8273	1.0147	1.4065	0.8727
16	**Dalarna**	0.8846	1.0479	1.1747	0.9573	1.0117	1.1336	1.3653	0.4691
17	**Gävleborg**	0.7186	1.6801	0.6224	0.3349	0.8778	2.1156	1.9937	1.1630
18	**Västernorrland**	0.9546	0.8847	0.4272	1.5008	1.2565	0.9242	1.4644	1.0300
19	**Jämtland**	1.0648	1.3413	0.5762	0.8647	1.1604	1.0163	1.3691	1.0440
20	**Västerbotten**	0.5295	1.5903	1.0503	0.4205	1.2091	0.9656	1.7347	1.0020
21	**Norrbotten**	1.2589	1.8902	0.5410	0.4229	1.5909	1.2375	1.2633	0.8967
	Sweden	1.2376	1.7442	-0.6796	-1.2150	-0.8326	-0.6361	-0.7303	0.3703

Table A.4. Estimates of $\Theta+\theta_{yr}$ and the standard deviation, σ_{θ}, of θ_{yr}.

		1992	1993	1994	1995	1996	1997	1998	s ?
1	**Stockholm**	2.9946	3.4445	3.2751	3.5811	3.4088	3.4318	3.3029	0.2766
2	**Uppsala**	3.3159	3.3292	3.2603	3.3818	3.5548	3.4268	3.3368	0.2718
3	**Södermanland**	3.1466	3.2045	2.9223	3.1423	3.3517	2.7743	3.5795	0.4206
4	**Östergötland**	3.4201	3.3858	3.4073	3.3690	3.3570	3.4959	3.2420	0.2371
5	**Jönköping**	3.1816	3.0706	3.1166	3.3387	3.2425	3.1588	3.1294	0.2186
6	**Kronoberg**	3.4049	3.4417	3.4214	3.5144	3.3327	3.3582	3.3245	0.2340
7	**Kalmar**	2.9629	2.9154	2.8976	2.8689	2.7777	2.8328	2.8548	0.1551
8	**Gotland**	2.9872	3.0244	2.9887	2.9982	2.9699	2.9829	2.9903	0.1382
9	**Blekinge**	3.1582	3.2945	3.2296	3.2979	3.2133	3.2732	3.2520	0.2252
10	**Skåne**	3.3919	3.1178	3.0866	3.2582	3.2340	3.2605	3.3316	0.1911
11	**Halland**	3.3400	3.4135	3.5140	3.5508	3.4667	3.3589	3.4185	0.1941
12	**Västra Götaland**	3.4915	3.2359	3.2690	3.4342	3.4747	3.3564	3.1498	0.2218
13	**Värmland**	3.1599	3.2724	3.3356	3.4403	3.2108	3.2411	3.1999	0.2288
14	**Örebro**	3.1793	3.0683	3.1449	3.3347	3.1570	3.1160	3.2601	0.2191
15	**Västmanland**	2.9594	2.9647	2.9624	3.2185	3.1519	2.9557	2.8932	0.2223
16	**Dalarna**	3.2645	3.3143	3.2965	3.1705	3.3578	3.1787	3.2310	0.1777
17	**Gävleborg**	3.1293	3.1169	3.1190	3.1106	3.1364	3.1768	3.1236	0.1217
18	**Västernorrland**	3.0170	2.9346	3.1361	3.2300	3.2004	3.0872	3.0353	0.1764
19	**Jämtland**	3.1263	3.0155	3.1503	3.1389	3.1968	3.1039	3.1080	0.1352
20	**Västerbotten**	2.9578	3.2212	4.0426	3.3263	3.4141	3.0502	3.1237	0.5443
21	**Norrbotten**	3.2008	3.2629	3.4210	3.3037	3.4926	3.2674	3.2862	0.2007
	Sweden	3.2364	3.1677	3.2546	3.4731	3.4644	3.2444	3.1715	0.1733

Table A.5. Estimates of b_{yr}.

		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
1	**Stockholm**	-1.6750	-1.4980	-0.8588	-0.2028	-1.3470	-1.9170	-1.1570	-1.5460
2	**Uppsala**	-1.6060	-1.0020	-0.7203	-0.9024	-1.1930	-1.4400	-0.6849	-1.4170
3	**Södermanland**	-2.3270	-1.6450	-1.5750	-0.5987	-1.7170	-2.4960	-0.4515	-2.0350
4	**Östergötland**	-0.7072	-1.2230	-1.4310	-1.1930	-1.2820	-1.8030	-0.8573	-0.5971
5	**Jönköping**	-1.6800	-0.7079	-0.3771	-0.9868	-0.8834	-0.9587	-1.1820	-0.1568
6	**Kronoberg**	-1.3820	-1.5960	-0.9501	-1.2790	-1.0520	-1.8580	-1.6410	-1.1290
7	**Kalmar**	-1.2330	-0.7411	-0.5438	-0.6957	-0.1894	-0.7272	-0.7975	-0.2833
8	**Gotland**	-1.6530	-1.1610	-0.7232	-0.5184	-0.3976	-0.5248	-1.5460	-0.5881
9	**Blekinge**	-0.8742	-2.8470	-3.6380	-1.7610	-1.4620	-1.8470	-1.4230	-1.7400
10	**Skåne**	-0.5720	-1.0890	-0.5796	-0.4665	-0.3088	-1.6990	-0.9341	-1.0550
11	**Halland**	-1.3210	-0.5562	-0.4421	-0.8710	-0.7984	-1.2700	-1.2510	-1.0010
12	**Västra Götaland**	-0.4299	-1.1110	-0.4744	-0.7368	-0.9422	-2.1270	-1.1930	-0.8319
13	**Värmland**	-2.4620	-1.5050	-1.0200	-0.0093	-0.4665	-0.9279	-1.3370	-0.6830
14	**Örebro**	-2.0890	-1.7250	-1.2680	-0.5314	-1.2190	-1.5100	-0.8705	-1.3880
15	**Västmanland**	-1.0340	-1.6830	-0.8108	-0.9380	-0.5781	-1.6380	-0.9725	-0.5242
16	**Dalarna**	-2.2850	-1.2980	-1.3230	-0.8699	-0.7858	-1.9050	-1.0790	-0.7149
17	**Gävleborg**	-2.1850	-1.7620	-0.4352	-1.2580	-1.8120	-2.1080	-0.6133	-0.8506
18	**Västernorrland**	-1.7270	-1.5320	-0.6474	-0.1985	-0.9703	-1.8290	-0.6650	-0.8335
19	**Jämtland**	-2.5230	-1.1850	-0.6755	-0.7944	-0.3461	-1.9350	-0.3991	-0.9080
20	**Västerbotten**	-1.5850	-1.6560	-0.3612	-1.0380	-0.9323	-1.6850	-1.6030	-0.5123
21	**Norrbotten**	-2.4060	-1.1200	-1.8440	-1.0790	-1.2160	-2.0210	-1.5400	-1.6120
	Sweden	-1.0100	-1.0970	-0.5656	-0.5066	-0.9533	-2.0720	-1.1360	-0.9215

Table A.6. Peak week for the counties in the period 1992-1998

	County	1992	1993	1994	1995	1996	1997	1998
1	**Stockholm**	29	33	32	34	33	33	32
2	**Uppsala**	32	32	32	33	34	33	33
3	**Södermanland**	31	31	29	31	32	28	34
4	**Östergötland**	33	33	33	33	32	34	31
5	**Jönköping**	31	30	30	32	31	31	31
6	**Kronoberg**	33	33	33	34	32	32	32
7	**Kalmar**	29	29	29	28	28	28	28
8	**Gotland**	29	30	29	29	29	29	29
9	**Blekinge**	31	32	31	32	31	32	32
10	**Skåne**	33	30	30	32	31	32	32
11	**Halland**	32	33	34	34	33	32	33
12	**Västra Götaland**	34	31	32	33	33	32	31
13	**Värmland**	31	32	32	33	31	31	31
14	**Örebro**	31	30	31	32	31	30	32
15	**Västmanland**	29	29	29	31	31	29	29
16	**Dalarna**	32	32	32	31	32	31	31
17	**Gävleborg**	31	30	30	30	31	31	30
18	**Västernorrland**	30	29	31	31	31	30	30
19	**Jämtland**	30	30	31	31	31	30	30
20	**Västerbotten**	29	31	38	32	33	30	30
21	**Norrbotten**	31	32	33	32	34	32	33

Table A.7. Ratio between high and low incidences:

	County	1992	1993	1994	1995	1996	1997	1998
1	Stockholm	6.4	4.0	4.2	6.5	9.1	9.5	9.3
2	Uppsala	3.9	3.5	4.3	6.0	5.5	5.9	6.4
3	Södermanland	10.5	10.8	7.9	8.9	10.5	9.3	10.6
4	Östergötland	2.9	8.0	6.4	4.6	9.0	5.0	5.7
5	Jönköping	4.0	4.8	4.1	4.6	4.8	5.2	4.9
6	Kronoberg	6.3	7.7	5.8	5.6	10.6	8.7	9.5
7	Kalmar	2.6	2.4	3.5	4.3	6.0	4.9	6.1
8	Gotland	4.8	6.7	4.8	7.2	5.0	5.8	5.8
9	Blekinge	14.0	13.6	9.8	14.6	11.8	14.2	14.7
10	Skåne	3.2	3.8	4.8	2.6	4.7	8.7	5.6
11	Halland	4.2	3.8	4.6	4.6	4.9	5.4	5.4
12	Västra Götaland	3.1	4.4	3.5	4.2	7.7	9.4	9,4
13	Värmland	7.0	5,8	5.9	3.5	5.1	6.3	10.2
14	Örebro	6.3	6.9	5.9	6.0	7.2	6.6	8.3
15	Västmanland	6.9	7.1	7.7	6.5	7.7	8.4	10.4
16	Dalarna	6.8	7.1	7.0	6.1	8.0	6.7	9.2
17	Gävleborg	10.7	6.2	7.5	8.5	13.6	10.4	10.8
18	Västernorrland	9.9	7.7	10.3	8.0	9.3	8.3	10.3
19	Jämtland	10.4	9.9	9.6	8.0	9.6	7.2	11.7
20	Västerbotten	7.0	8.7	8.4	10.0	7.5	7.8	11.0
21	Norrbotten	11.6	13.7	10.9	10.6	11.1	11.4	15.4

Table A.8. Week when the 2-times-base incidence period starts.

	County	1992	1993	1994	1995	1996	1997	1998
1	Stockholm	17	18	20	20	15	16	13
2	Uppsala	19	21	18	15	19	19	18
3	Södermanland	14	14	13	13	15	13	19
4	Östergötland	21	15	16	16	15	17	16
5	Jönköping	21	19	19	21	20	20	19
6	Kronoberg	17	16	18	17	15	16	15
7	Kalmar	24	24	22	21	20	21	20
8	Gotland	21	19	22	20	22	21	21
9	Blekinge	16	6	16	12	12	14	13
10	Skåne	23	20	18	23	19	18	20
11	Halland	19	22	21	21	21	20	20
12	Västra Götaland	23	19	20	21	17	17	15
13	Värmland	21	22	23	25	23	21	20
14	Örebro	15	15	15	17	15	15	16
15	Västmanland	20	20	19	21	20	19	19
16	Dalarna	17	18	18	17	18	17	18
17	Gävleborg	17	21	17	16	17	21	20
18	Västernorrland	22	21	22	24	23	22	22
19	Jämtland	24	23	24	24	24	24	24
20	Västerbotten	18	21	27	20	23	19	20
21	Norrbotten	18	20	16	14	21	18	18