Research Report
Statistical Research Unit
Department of Economics
Göteborg University
Sweden

Explorative analysis of spatial aspects on the Swedish influenza data.

David Bock, Kjell Pettersson

Research Report 2007:10
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-786 12 74	Nat: 031-786 00 00	http://www.statistics.gu.se/
P.O. Box 640	Int: +46 31 786 12 74	Int: +46 31 786 00 00	
SE 405 30 Göteborg			
Sweden			

Summary

The spatial aspects on the Swedish influenza data are analyzed. During the influenza period, reports on laboratory diagnosed cases and influenza-like-illness are obtained from viral and microbiological laboratories and from sentinel physicians, respectively, in different regions of Sweden. Information about the spatio-temporal patterns might give insight in the way the influenza spreads over Sweden. It might also be used in automated surveillance systems for outbreak and peak detection of the influenza. We describe the regional patterns in Swedish influenza data in different ways. Several natural hypotheses about geographical patterns are examined but can not be verified as consistent over the years. However, we find that, for a group of large cities, the outbreak of the influenza occurs at least four weeks earlier than for the rest of Sweden. The possibilities to utilize this in surveillance systems are briefly discussed.

Innehåll

1 INTRODUCTION

Each year the ordinary epidemic influenza causes considerable suffering and costs for individuals as well as for society. In order to take timely preventive actions such as allocating medical resources, it is important to detect the outbreak of the annual influenza epidemic. For actions to be effective, systems for outbreak detection should be timely and detect the outbreak early without having too many false alarms.

It may also be important to detect the peak of the influenza epidemic. If the ordinary influenza has reached its peak, an additional increase in the number of cases of influenza-like illness could indicate the outbreak of another infectious disease with the same symptoms.

During the influenza season, data on the number of laboratory diagnosed influenza cases (LD) and the number of patients with influenza-like illnesses (ILI) are provided on a weekly basis from different laboratories and different sentinel physicians, respectively, in different parts of Sweden. The data are reported to the Swedish Institute for Infectious Disease Control (SMI). Information on the Swedish influenza incidence is found in the publications by the Swedish Institute for Infectious Disease Control, e.g. Linde et al. (2004) and in international publishing such as Ganestam et al. (2003). An exploratory analysis of ILI and LD on a national level where made in Andersson et al. (2006). Detection of peaks in LD data was considered in Bock et al. (2006).

It is of great value to consider spatio-temporal variations in a surveillance system since influenza might appear in different areas at different time points. There are several examples in the literature. One example is the winter of 1993-1994 when the epidemic started in Scotland and spread south to other European countries via England and France (Fleming and Cohen (1996)). A considerable time lag was evident. Such information can be used in a surveillance system to make it more effective, see e.g. Järpe (2000) and Järpe (2001). In addition, the time lag can be used to gain time for preventive actions.

In this report we examine exploratory the spatial aspects of the Swedish influenza incidence as measured by ILI and LD. We study the development of outbreak of the influenza in Sweden. We also investigate the spatio-temporal aspect of the influenza peak (when the incidence starts to decrease). Influenza-data from other countries might also be valuable in this analysis. However, only Swedish data are considered here. The emphasis is on exploratory analysis. Even though the motivation of this analysis is to give information necessary for the construction of surveillance systems, the actual construction is only briefly discussed in this report.

The plan of this paper is as follows. In section 2 the Swedish data on influenza is discussed. In section 3 and 4 we analyze ILI and LD, respectively. Some concluding remarks are given in section 5.

2 SWEDISH DATA ON INFLUENZA

Two different types of data are analyzed in this report: weekly data on ILI and LD in Sweden. The reporting systems are described in e.g. Ganestam et al. (2003), Andersson et al. (2006) and in the annual reports from The National Influenza Reference at SMI (www.smittskyddsinstitutet.se). Data on e.g. mortality, records on sick leave and reason for being absent from work might then be valuable but are not analyzed here. Covariates, such as gender and age of patients with ILI or those with laboratory confirmed influenza were not available. The data on ILI and LD are briefly described below.

2.1 Influenza-like-illness

Data on ILI are collected from a number of voluntarily participating sentinel physicians representing about 2% of the Swedish general practitioners. The sentinel physicians report the number of patient visits (#PAT) and the number of patients with ILI (#ILI). In this paper we analyze data from five periods, 00_01, 01_02, 02_03, 03_04 and 04_05.

One estimate of the influenza incidence is the percentage of the visiting patients showing influenza-like-illness (%ILI). The total number of reported patient visits varies a lot between years as well as between weeks as seen in Figure 1.

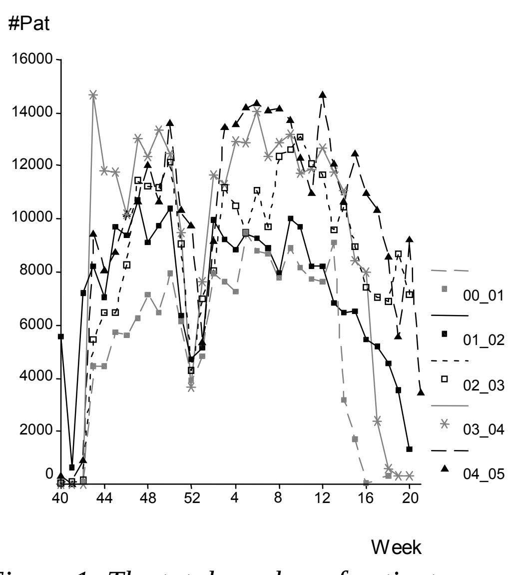


Figure 1: The total number of patients reported by the sentinel physicians.

Few cases are reported around Christmas. The physicians' expectation of whether the outbreak or peak has occurred might influence their motivation to report. Due to this variability the variances of the estimates of the incidence will vary much. The %ILI data should therefore be interpreted with care since it might be influenced by different time dependent effects, such as lack of motivation, as pointed out by Andersson et al. (2006).

In the data available for the periods 03_04 and 04_05 the records start at week 40 with several consecutive weeks of zero values on #PAT. We interpreted these zero records as absence of reports, i.e. for these two periods the start of the reporting was defined as the first week with non-zero reports on #PAT. For 00_01 to 02_03 there are no records on weeks with zero patients in the beginning. Instead the starting week of reporting varies. The resulting weeks with the first and the last report during the influenza seasons are given in Table 1 and we see that the start and the end of the reporting vary slightly between regions. Most of the regions report their first non-zero value of #PAT in week 43 and their last in week 20 the following year.

Table 1. The weeks of the first and last report on ILI from the sentinel physicians.

	00_01		***01_02***		***02_03***		***03_04***		***04_05***	
Region	***Start***	***End***	***Start***	***End***	***Start***	***End***	***Start***	***End***	***Start***	***End***
Blekinge	6	14	43	19	44	20	43	16	43	19
Dalarna	43	15	40	19	43	20	44	16	40	19
Gävle	43	18	40	20	43	20	43	17	40	20
Halland	43	15	45	19	46	20	43	16	46	19
Jönköping	43	16	40	14	46	20	43	15	10	20
Jämtland	-	-	40	19	43	20	43	16	42	19
Kalmar	43	13	42	12	50	19	43	16	46	20
Kronoberg	43	14	43	19	43	20	43	16	40	19
Norrbotten	43	14	43	20	43	17	43	17	43	20
Skåne	43	16	40	20	43	20	43	18	40	20
Stockholm	46	13	40	19	43	20	43	16	43	20
Södermanland	43	13	41	18	44	20	44	17	42	20
Uppsala	44	12	40	19	42	20	43	17	42	20
Västra Götaland	42	15	40	20	40	20	43	20	42	20
Västernorrland	46	12	41	16	43	13	43	15	40	19
Värmland	43	14	40	19	43	20	43	15	40	19
Västmanland	43	15	40	20	43	20	43	18	44	20
Västerbotten	46	14	40	19	43	20	43	16	43	20
Örebro	43	13	41	19	43	19	43	17	45	17
Östergötland	45	13	40	19	46	15	44	17	43	19

Since the influenza starts during late autumn or winter, it is the incidence during the autumn that should represent the baseline process. However, due to the sparseness of the available data estimating the baseline is difficult. Graphs of %ILI for the whole Sweden are shown in Figure 2.

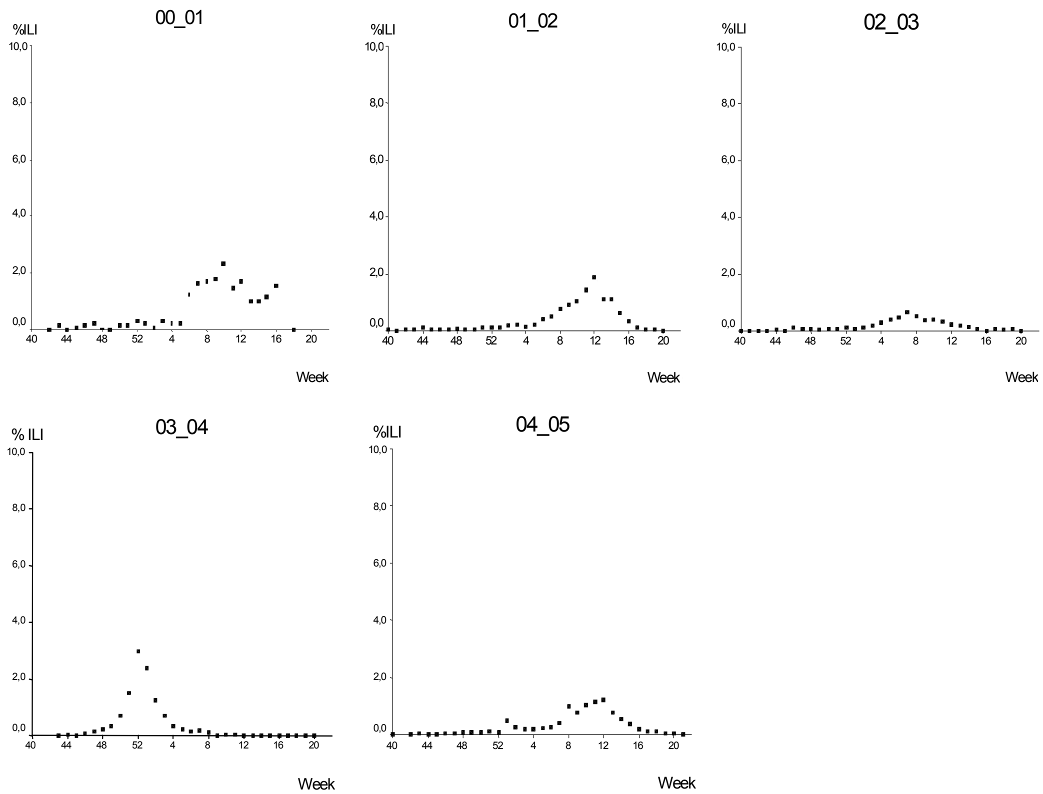


Figure 2. %ILI for the whole of Sweden.

2.2 Laboratory diagnosed cases

Data on LD are reported from five viral laboratories (at university hospitals and at SMI) and a number of microbiological laboratories. The data analyzed here are from six periods, 99_00, 00_01, 01_02, 02_03, 03_04 and 04_05. There are three virus types of influenza (A, B and C) which all belong to the group ortomyxovirus. It is mainly A and B that give rise to the typical influenza disease (see Andersson et al. (1999)). In the laboratories each influenza case is typed. In our analyses LD is defined as the sum of A and B cases. Graphs of LD for the whole of Sweden are shown in Figure 3.

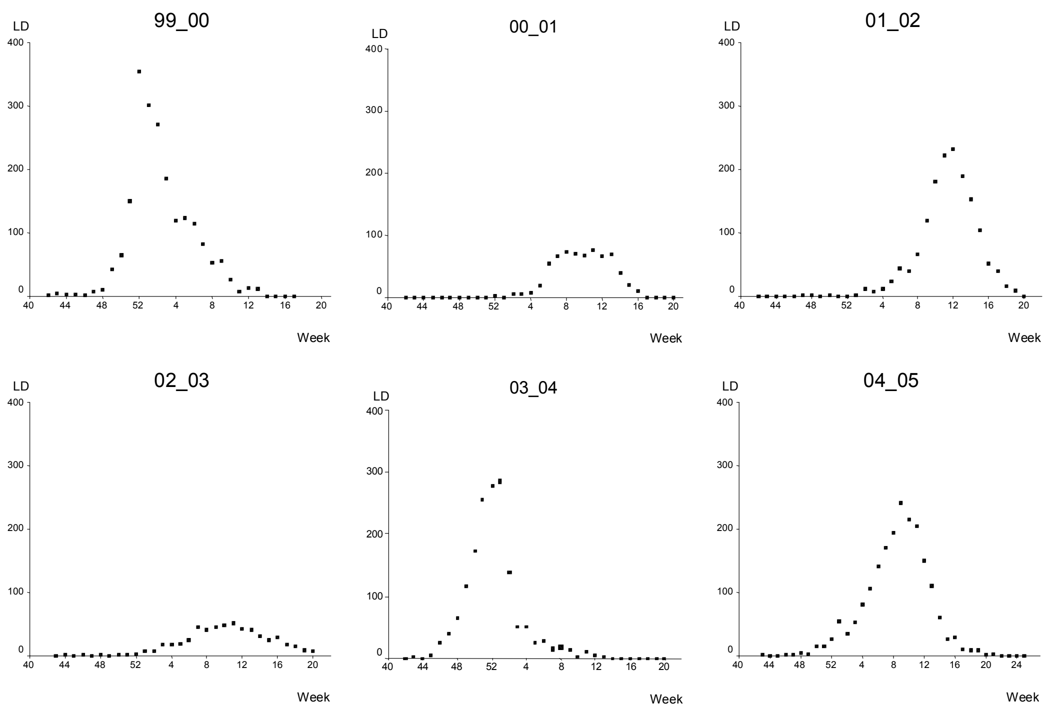


Figure 3. LD for the whole of Sweden.

The number of laboratories available varies slightly between the periods. In 99_00 and 00_01 there are 17 laboratories. In 01_02 there are 21 laboratories and in 02_03 there are 23 laboratories. There are 24 in 03_04 and 04_05. There is in general one laboratory in each larger city. In the Stockholm region there are two laboratories at Huddinge University Hospital (HS) and Karolinska University Hospital (KS). The weeks of the first and the last report during the influence seasons are given in Table 2.

Table 2. The weeks of the first and last report on LD from the laboratories.

	99_00		00_01		01_02		02_03		03_04		04_05	
Laboratory	**Start**	**End**	**Start**	**End**	**Start**	**End**	**Start**	**End**	**Start**	**End**	**Start**	**End**
Borås	-	-	-	-	-	-	-	-	46	6	53	16
Eskilstuna	-	-	-	-	-	-	3	15	47	3	5	9
Falun	50	12	5	12	5	15	1	20	48	5	1	18
Gävle	-	-	-	-	6	13	5	16	43	3	3	14
Göteborg	49	10	2	12	2	16	46	17	46	6	48	16
Halmstad	49	13	6	11	2	17	5	16	47	6	49	14
HS	43	13	5	12	48	17	1	20	43	9	47	21
Jönköping	-	-	-	-	-	-	-	-	51	2	11	11
Kalmar	49	13	8	12	4	16	11	19	45	8	3	15
Karlskrona	-	-	-	-	4	17	4	16	47	4	6	13
Karlstad	46	12	7	12	2	16	3	10	48	8	51	16
KS	43	13	4	12	47	17	48	20	45	9	47	19
Linköping	44	6	46	12	6	15	7	19	45	52	50	14
Luleå	51	8	-	-	-	-	52	8	50	7	3	15
Malmö	43	11	52	12	50	17	3	20	48	9	48	16
Skövde	48	13	2	12	3	16	44	20	45	8	53	21
Sundsvall	-	-	9	10	2	17	8	14	48	9	3	14
Uddevalla	51	13	4	10	4	14	7	17	47	3	49	16
Umeå	43	13	46	12	3	17	52	18	47	9	50	21
Uppsala	48	9	2	12	2	16	3	15	48	6	43	15
Västerås	1	13	51	12	10	13	8	10	49	3	51	16
Växjö	52	10	6	12	4	16	6	12	47	1	12	12
Örebro	50	13	52	12	4	17	6	16	46	8	50	18
Östersund	-	-	-	-	7	17	-	-	8	8	1	18

The week the reporting starts and ends varies considerably between the laboratories, much more than for ILI. The numerous empty records in the table are due to the variation in the number of participating laboratories each period. The periods 99_00 and 00_01 are the only ones where a few laboratories (Linköping in 99_00 and Linköping, Umeå, Västerås and Örebro in 00_01) report zero cases in the beginning and it might be that the laboratories wait to start sending reports until they have their first confirmed case of influenza. The first confirmed case is consequently very important and could be interpreted as an outbreak. The baseline level of LD would then be zero cases.

LD is probably a more reliable variable compared to %ILI for the reasons pointed out in the previous section. That was also concluded by Andersson et al. (2006) who considered %ILI to be more appropriate for the non-epidemic period and outbreak since the first report of an LD-case often comes later than the first reported ILI-case. LD was considered more appropriate for the period near the peak. In this paper, however, both %ILI and LD will be investigated for the non-epidemic period and outbreak as well as the peak and decline.

3 INFLUENZA-LIKE-ILLNESS

3.1 Non-epidemic period and outbreak

In this section we study the variation between the regions in the time-point when the influenza incidence, as measured by %ILI, starts to increase. The collection of Swedish influenza data during non-epidemic periods is very sparse and therefore, we have very little information about the statistical properties of the baseline. In order to determine the time-points of the increase we study the time until the cumulative sum $C_t = \sum_{i=1}^{t} \%ILI(i)$ exceeds a threshold value, c, where c is a constant independent of t. The cumulative sum is used in order to smooth the incidence. This is made for each period and region and the different values of c = {0.10, 0.25, 0.5, 1}. For illustration, the incidence and the cumulative incidence are plotted in Figure 4 for some regions.

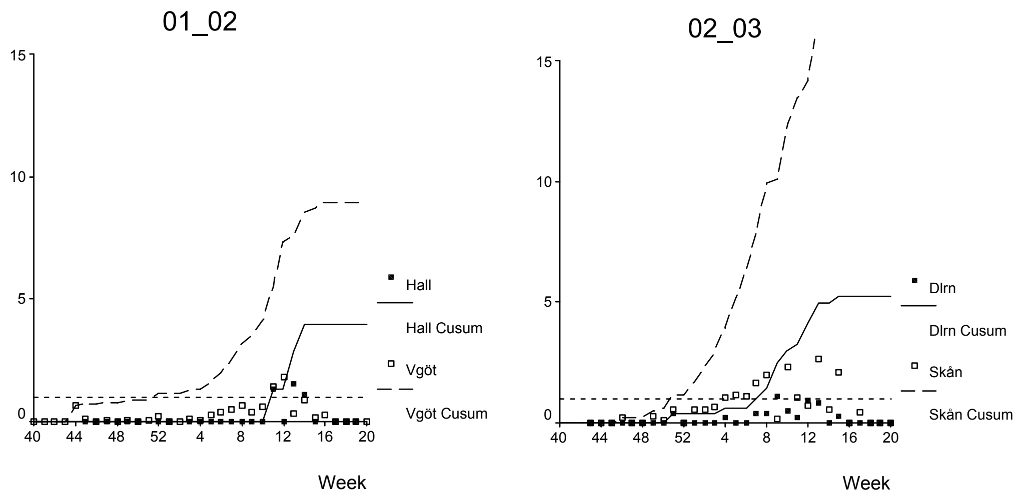


Figure 4. %ILI and cumulative %ILI for the regions Halland (Hall), Västra Götaland (Vgöt), Dalarna (Dlrn) and Skåne (Skån). The value c=1 is marked with a dashed horizontal line. The number of weeks until the cumulative incidences exceed c is 24, 13, 20 and 12, respectively.

The first time when $C_t > c$ should not necessarily be interpreted as the time of the true outbreak but a simple way to characterize the evolution of the epidemic. As was mentioned in section 2.1 the week for the first report varies slightly between the regions. In order to number the weeks consecutively, week 40 is used as starting week for all the periods. Results are given in Table 3 for c=0.5.

Table 3. The number of weeks until the cumulative sum of %ILI, C_t, exceeds the limit c=0.5 . Week 40 is used as the starting week. The median and range over the periods are reported.

Region	**00_01**	**01_02**	**02_03**	**03_04**	**04_05**	**Median**	**Range**
Blekinge	17	7	16	8	7	8	10
Dalarna	2	4	17	10	7	7	15
Gävle	12	14	20	8	18	14	12
Halland	11	24	17	9	20	17	15
Jönköping	9	14	20	20	24	20	15
Jämtland	-	7	23	12	21	16,5	16
Kalmar	5	20	21	13	19	19	16
Kronoberg	9	19	16	9	16	16	10
Norrbotten	17	19	11	9	14	14	10
Skåne	4	9	10	10	12	10	8
Stockholm	9	7	7	7	12	7	5
Södermanland	18	3	17	10	11	11	15
Uppsala	10	8	10	11	15	10	7
Västra Götaland	17	5	16	9	11	11	12
Västernorrland	14	17	11	4	14	14	13
Värmland	11	17	22	11	13	13	11
Västmanland	15	19	20	10	15	15	10
Västerbotten	5	4	7	8	13	7	9
Örebro	21	24	21	11	33	21	22
Östergötland	20	27	26	14	20	20	13

One might expect a spread of influenza in some well-defined directions such as from the south to the north or from the west to the east. It also seems reasonable to assume that there is a spread from one region to geographically close regions. We found, however, no evidence for such patterns in the data studied. Another hypothesis is that the large cities are earlier than other regions when it comes to an outbreak of influenza. We found a time-lag between two groups of regions in the increase of the incidence. Group 1 comprises the following regions Stockholm-Uppsala, Göteborg and Malmö, that is regions with large cities. Group 2 includes the other 16 regions. The within-group median number of weeks until C_t exceeds the threshold c is reported in Table 4 for each period and value of c.

Table 4. Within-group median number of weeks until the cumulative sum of %ILI, C_t, exceeds the limit c for each period and value of c. Week 40 is used as the starting week.

	00_01		**01_02**		**02_03**		**03_04**		**04_05**	
Group	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
c=1	13.5	14	13,5	19,5	15,5	20	11	11	14,5	19
c=0.50	9.5	12	7,5	17	10	18,5	9,5	10	12	15,5
c=0.25	8	11	5	15	9	17	8,5	9	11,5	14
c=0.10	6.5	11	4	10,5	7	16,5	8	8,5	8,5	13,5

Table 5. Median number of weeks until the cumulative sum of %ILI, C_t, exceeds the limit c for each group over all periods. Week 40 is used as starting week.

Group	c=1	c=0.5	c=0.25	c=0.10
1	12,5	10	9	7
2	17	14	13	11

The median times over the years are shown in Table 5 for the two groups. The difference between the median time-points in the two groups is approximately four weeks. The variation between the different periods as measured by the range has a tendency to be smaller in group 1 than in group 2. The conclusion is that the influenza starts in large cities.

3.2 Peak

The number of weeks until the maximum value of %ILI is reached is investigated here. The maximum value can be interpreted as an estimate of the peak of the incidence. The results are shown in Table 6.

Table 6. The number of weeks until the maximum value of %ILI. Week 40 is used as starting week. The median and range over the periods are reported. The regions are sorted with respect to the median.

Region	00_01	01_02	02_03	03_04	04_05	Median	Range
Västerbotten	15	23	19	15	14	15	9
Norrbotten	21	25	16	14	14	16	11
Stockholm	20	25	18	13	14	18	12
Uppsala	20	26	19	13	17	19	13
Blekinge	19	22	20	14	25	20	11
Gävle	20	22	20	15	25	20	10
Södermanland	21	28	20	14	16	20	14
Västra Götaland	18	25	20	13	24	20	12
Västmanland	20	24	20	13	30	20	17
Halland	21	26	19	14	25	21	12
Kronoberg	21	25	21	15	27	21	12
Örebro	21	25	22	16	19	21	9
Dalarna	20	25	22	14	26	22	12
Jönköping	22	24	22	20	24	22	4
Västernorrland	17	22	24	10	26	22	16
Östergötland	22	27	22	14	25	22	13
Värmland	21	23	24	13	24	23	11
Jämtland	-	25	24	12	23	23,5	13
Kalmar	23	24	24	13	25	24	12
Skåne	19	26	26	14	26	26	12

Group 1 consisting of larger cities (see Section 3.1) has peaks occurring approximately 1 week earlier than group 2, which consists of the 16 other regions. An exception is the period 01_02 where the reverse holds (see Table 7).

Table 7.LD: Within-group median number of weeks number of weeks until the maximum value. Week 40 is used as starting week.

Group	*00_01*	*01_02*	*02_03*	*03_04*	*04_05*
1	19,5	25,5	19,5	13	20,5
2	21	24,5	21,5	14	25

The relation between northern regions (Jämtland, Norbotten, Västerbotten and Västernorrland) and the others were also investigated. The former regions were also approximately 1 week earlier than the others except for the period 02_03. The conclusion is that patterns found in the data available are rather weak.

Table 8. The height of the peak of %ILI as measured by the maximum value of %ILI. The median and range over the periods are reported. The regions are sorted with respect to the median.

Region	*00_01*	*01_02*	*02_03*	*03_04*	*04_05*	*Median*	*Range*
Kalmar	1,20	3,45	0,40	0,45	1,13	1,13	3,05
Östergötland	1,43	0,31	0,29	1,50	1,21	1,21	1,21
Gävle	3,26	1,26	0,43	1,19	1,40	1,26	2,83
Örebro	1,74	1,24	1,83	1,40	0,23	1,40	1,60
Västmanland	3,12	1,69	0,29	3,82	1,40	1,69	3,53
Jämtland	-	1,90	1,99	2,30	0,77	1,95	1,53
Västra Götaland	3,55	1,80	0,39	1,95	1,99	1,95	3,16
Uppsala	3,83	2,35	1,45	5,30	1,50	2,35	3,85
Halland	3,26	1,54	2,68	11,71	2,40	2,68	10,17
Värmland	3,55	2,77	0,42	1,42	4,50	2,77	4,08
Södermanland	3,26	2,94	2,78	6,90	1,77	2,94	5,13
Kronoberg	3,49	3,02	1,52	1,25	3,45	3,02	2,24
Västerbotten	4,48	9,73	1,65	3,18	3,08	3,18	8,08
Stockholm	3,59	3,23	5,13	7,01	1,38	3,59	5,63
Jönköping	4,02	10,53	3,85	3,85	7,14	4,02	6,68
Skåne	11,11	4,32	2,63	11,20	4,68	4,68	8,57
Blekinge	2,94	6,25	5,26	16,90	13,04	6,25	13,96
Västernorrland	4,17	5,06	6,25	6,78	8,89	6,25	4,72
Dalarna	19,35	4,66	1,08	10,29	6,77	6,77	18,27
Norrbotten	7,19	28,00	100,00	6,58	1,89	7,19	98,11

In Table 8 we see that Norrbotten has some very large maxima. However, there are few patient visits reported and hence the variance is large.

Linde and Ekdahl (2003) point out that influenza periods that start early cause more severe epidemics compared to periods that start late. Andersson et al. (2006) demonstrate that the time of the outbreak can be used as a predictor for the height of the peak of the influenza. For each period the correlations between the time of the outbreak as defined in section 3.1 and the height of the peak are negative with the exception of the period 01_02. From the data available, there is hence support for the conclusion that early outbreaks are followed by high peaks.

3.3 Further comparison of the group of large cities with other regions.

In section 3.1 it was found that the difference between median time-points for the increases in %ILI between group 1 (the large cities) and group 2 (the other regions) was approximately 4 weeks, i.e. group 1 can be considered as a leading indicator for the incidence of the other group as well as for the whole of Sweden. We will now give a further description of the two groups and the whole of Sweden.

3.3.1 Parametric models of the expected incidence

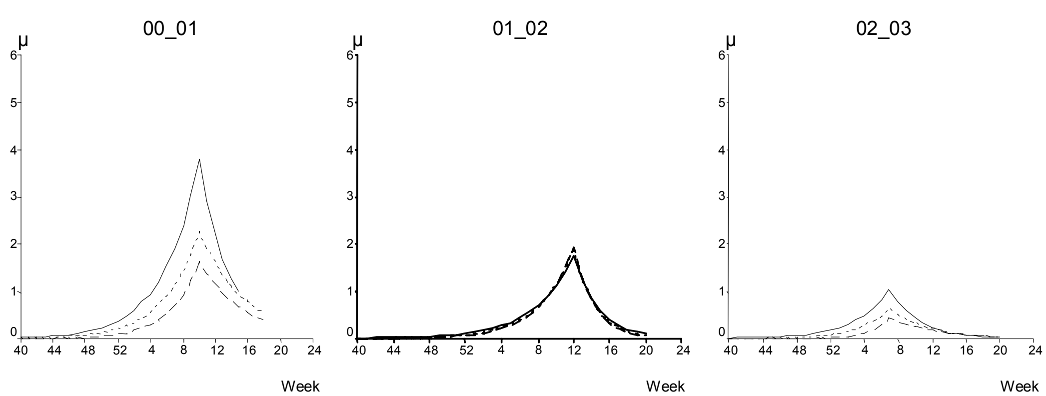
Andersson et al. (2006) discussed several simple parametric curve forms for the expected incidence, μ, for the whole of Sweden. It was concluded that a piecewise exponential curve,

$$\mu(t \mid \tau) = \begin{cases} \beta_0 \cdot \exp(\beta_1 \cdot t), & t < \tau \\ \beta_0 \cdot \exp(\beta_1 \cdot (\tau\text{-}1) + \beta_2 \cdot (t\text{-}\tau + 1)), & t \geq \tau \end{cases},$$

where t =1, 2,… is time and τ is the first time of the decline, resulted in a reasonable fit. The parameters of this curve are estimated for the two groups and for the whole of Sweden. The parameter values are given in Table 9. The estimated curves are shown in Figure 5.

Table 9. %ILI: Estimated parameter values for the exponential curve.

	Group 1			Group 2			Whole		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
00_01	0,01947	0,22926	-0,27	0,00267	0,278672	-0,17379	0,010877	0,232178	-0,17595
01_02	0,00601	0,22752	-0,3583	0,00222	0,270795	-0,4355	0,003066	0,256784	-0,4127
02_03	0,00785	0,24386	-0,2822	9,9E-05	0,420005	-0,17554	0,002369	0,280653	-0,2171
03_04	0,00016	0,79919	-0,7203	0,00471	0,451549	-0,65328	0,000249	0,726576	-0,453
04_05	0,00828	0,21771	-0,3619	0,00572	0,232571	-0,42521	0,013035	0,18466	-0,47303



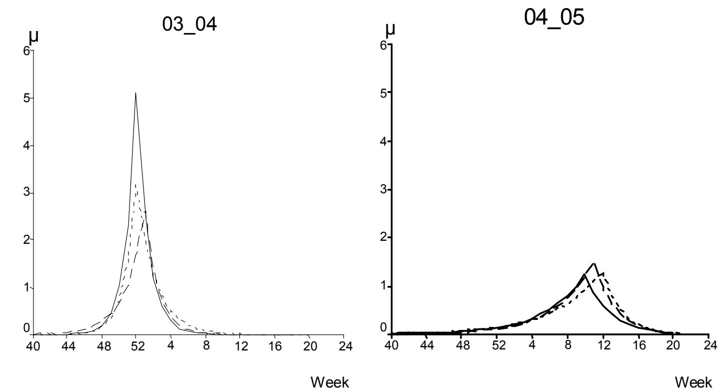


Figure 5. The exponential curves estimated for %ILI. Group 1 (——), group 2 (− − −), the whole of Sweden (- - - -).

The leading property of group 1 is seen in Figure 5 for the periods 00_01, 02_03 but hard to see for the other periods since the peaks dominate the visual impression and the leading property was demonstrated for the outbreak. It can be seen that group 1 has higher peaks than group 2 during the periods 00_01, 02_03 and 03_04.

The estimated curve can be used for prediction of the outbreak and the peak. The curve could either be based on only the regions in group 1, the whole of Sweden or some kind of combination of the groups, as will be further discussed in section 5.

The residual variances from the estimated exponential curves are shown in Table 10. In all the five periods, group 1 has a larger residual variance than the whole of Sweden.

Table 10. %ILI: Residual variances from the estimated exponential curves.

	00_01	***01_02***	***02_03***	***03_04***	***04_05***
	σ^2	σ^2	σ^2	σ^2	σ^2
Group 1	0,2553	0,0183	0,0069	0,0099	0,0155
Group 2	0,082	0,0143	0,003	0,0102	0,0099
Whole	0,0805	0,0067	0,0016	0,0086	0,0114

The autocorrelations of the residuals are shown in Table 11. In two periods group 1 has (in absolute values) a larger first order autocorrelation compared to the whole of Sweden. Group 1 has larger second order autocorrelations in four periods. However, none of the autocorrelations are statistically significant.

Table 11. %ILI: First and second order autocorrelations of the residuals from the estimated exponential model.

	00_01		***01_02***		***02_03***		***03_04***		***04_05***	
	ρ(1)	ρ(2)	ρ(1)	ρ(2)	ρ(1)	ρ(2)	ρ(1)	ρ(2)	ρ(1)	ρ(2)
Group 1	0,338	-0,157	0,165	0,110	-0,057	-0,064	0,095	-0,245	-0,206	-0,165
Group 2	-0,219	-0,121	0,284	-0,280	0,005	0,042	-0,023	-0,006	0,074	-0,074
Whole	0,194	-0,256	0,182	-0,013	0,308	0,035	-0,111	-0,149	-0,092	-0,043

3.3.2 Non-parametric estimates of the expected incidence

When parametric assumptions are to be avoided unimodal regression can be used to estimate the expected incidence using only order restrictions. We know that μ has a peak in every influenza period but the shape of the curve (the height and the time of the peak), varies between periods. It is therefore difficult to use a parametric model for μ. Here only the information of unimodality is used:

$$\mu(1) \le \mu(2) \le \ldots \le \mu(j-1) \le \mu_{max} \text{ and } \mu_{max} \ge \mu(j) \ge \mu(j+1) \ge \ldots \ge \mu(t),$$

where μ_{max} is the peak value (not necessarily observable). In the unimodal regression, the aim is to estimate μ as a function of time t, under order restrictions. The estimation technique is described by Frisén (1986).

Andersson et al. (2006) estimated the curve for the whole of Sweden with this technique. In Figure 6 the non-parametric estimates are given also for group 1 (the large cities) and the other regions.

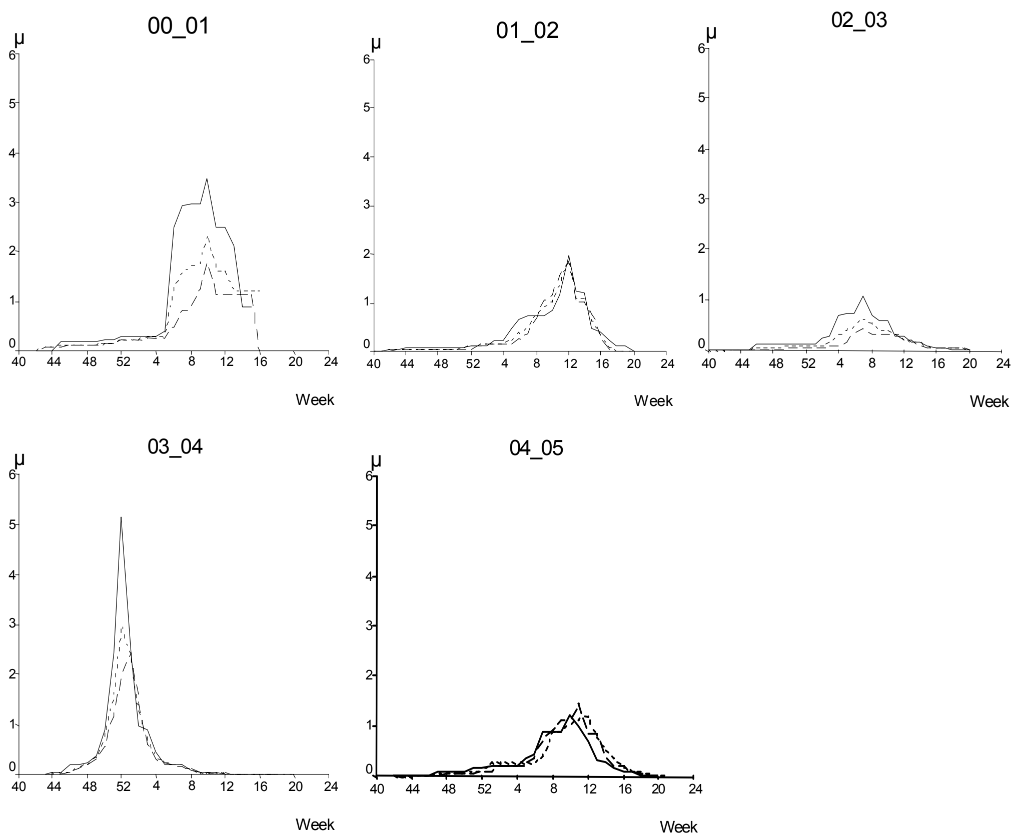


Figure 6. The estimated values of the incidence based on %ILI by the unimodal regression. Group 1 (——), group 2 (– – –), the whole of Sweden (- - - -).

The leading property of group 1 is seen in Figure 6 for the periods 00_01, 01_02 and 02_03 but is hard to see for the other periods since the peaks dominate the visual impression and the leading property was demonstrated for the outbreak.

4 LABORATORY DIAGNOSED CASES

4.1 Non-epidemic period and outbreak

We study the number of weeks until the first laboratory verified case of influenza, i.e. the number of weeks until LD(t) ≥ 1. Week 43 is used as starting week. The results are shown in Table 12.

Table 12. The number of weeks until the first LD case. Week 43 is used as starting week.The median and range over the periods are reported. The regions are sorted with respect to the median.

Laboratory	*99_00*	*00_01*	*01_02*	*02_03*	*03_04*	*04_05*	*Median*	*Range*
KS	1	13	5	6	3	5	5	12
HS	1	16	6	12	1	5	5,5	15
Göteborg	7	13	13	4	4	4	5,5	9
Borås	-	-	-		4	11	7,5	7
Malmö	1	10	16	14	6	6	8	15
Skövde	6	13	14	2	3	11	8,5	12
Umeå	1	16	14	10	5	8	9	15
Örebro	8	10	15	17	4	8	9	13
Luleå	9	-	-	10	8	14	9,5	6
Uppsala	6	13	13	14	6	1	9,5	13
Halmstad	7	17	13	16	5	7	10	12
Karlstad	4	18	13	14	6	9	11	14
Uddevalla	9	15	15	18	5	7	12	13
Falun	8	16	16	12	6	12	12	10
Linköping	2	17	17	18	3	8	12,5	16
Sundsvall	-	-	13	19	6	14	13,5	13
Eskilstuna	-	-		14	5	16	14	11
Kalmar	7	19	15	22	3	14	14,5	19
Gävle	-	-	17	16	1	14	15	16
Karlskrona	-	-	15	15	5	17	15	12
Västerås	12	22	21	19	7	9	15,5	15
Jönköping	-	-		-	9	22	15,5	13
Växjö	10	17	15	17	5	23	16	18
Östersund	-	-	18	-	19	12	18	7

The dispersion, as measured by the range, is large and several laboratories have no reporting for some periods. Group 1, with the larger cities Stockholm-Uppsala, Göteborg and Malmö, tend to have the first case of LD earlier than the other regions. The median times of the number of weeks until the first LD case are given in Table 13.

Table 13.LD: Within-group median number of weeks until the first LD case Week 43 is used as starting week.

Group	***99_00***	***00_01***	***01_02***	***02_03***	***03_04***	***04_05***
1	1	13	13	12	4	5
2	7,5	17	15	16	5	12

The median values for group 1, group 2 and the whole of Sweden over all periods are 6, 12.5 and 11, respectively. The conclusion is that the first LD case is reported approximately six weeks earlier in large cities than in the rest of the country. This does not automatically prove that the influenza starts in the large cities since the regions are of different size. However, it supports this conclusion in Section 3.1 based on %ILI data.

4.2 Peak

In this section the time of the maximum value of LD is studied. The results are shown in Table 14.

Table 14. The number of weeks until the maximum value of LD. Week 43 is used as starting week. The median and range over the periods are reported. The regions are sorted with respect to the median.

Laboratory	***99_00***	***00_01***	***01_02***	***02_03***	***03_04***	***04_05***	***Median***	***Range***
Borås	-	-	-	-	4	20	12	16
Luleå	12			19	12	15	13,5	7
Jönköping	-	-	-	-	9	22	15,5	13
Eskilstuna	-	-	-	26	5	16	16	21
Göteborg	12	17	20	19	9	17	17	11
Skövde	13	18	23	17	12	21	17,5	11
Umeå	12	22	23	21	12	15	18	11
HS	10	19	23	18	10	20	18,5	13
Uddevalla	13	18	25	20	7	19	18,5	18
Gävle	-	-	20	17	9	20	18,5	11
KS	10	18	22	21	10	20	19	12
Östersund	-	-	26	-	19	18	19	8
Karlskrona	-	-	26	18	5	20	19	21
Uppsala	13	20	21	22	12	19	19,5	10
Linköping	10	21	22	24	8	18	19,5	16
Karlstad	13	21	22	18	12	21	19,5	10
Örebro	17	21	21	18	10	21	19,5	11
Falun	10	23	22	19	10	20	19,5	13
Växjö	18	20	22	20	12	23	20	11
Malmö	10	18	24	23	12	22	20	14
Västerås	13	22	22	21	10	19	20	12
Halmstad	10	19	22	27	8	23	20,5	19
Kalmar	13	22	24	22	9	20	21	15
Sundsvall	-	-	23	20	6	22	21	17

No large difference in time of the peak between group 1 (larger cities) and 2 (other regions) appeared to be present, neither was there a difference between northern regions and the rest of Sweden. The conclusions regarding the time of the peak is that we did not find evidence for any pattern.

Table 15. The height of the peak of LD as measured by the maximum value of LD. The median and range over the periods are reported. The regions are sorted with respect to the median.

Laboratory	***99_00***	***00_01***	***01_02***	***02_03***	***03_04***	***04_05***	***Median***	***Range***
Karlskrona	-	-	2	1	1	4	1,5	3
Gävle	-	-	1	1	3	4	2	3
Östersund	-	-	3	0	1	3	2	3
Luleå	6	-	-	1	4	3	3,5	5
Växjö	7	5	12	3	3	1	4	11
Västerås	4	1	4	1	10	6	4	9
Eskilstuna	-	-	-	1	4	6	4	5
Borås	-	-	-	-	7	4	5,5	3
Jönköping	-	-	-	-	6	6	6	0
Linköping	10	2	8	4	7	7	7	8
Karlstad	30	4	8	3	8	11	8	27
Sundsvall	-	-	7	1	11	9	8	10
Kalmar	10	1	7	3	10	18	8,5	17
Halmstad	22	5	9	3	8	11	8,5	19
Uddevalla	18	5	13	3	11	8	9,5	15
Göteborg	21	13	8	5	16	9	11	16
Uppsala	17	7	15	4	10	22	12,5	18
Örebro	23	5	16	4	26	17	16,5	22
Falun	15	5	26	7	42	22	18,5	37
HS	59	17	28	10	37	22	25	49
Skövde	23	13	33	6	29	40	26	34
Umeå	49	26	43	15	38	24	32	34
Malmö	36	7	29	13	44	55	32,5	48
KS	70	20	35	11	54	33	34	59

The heights of the peaks of group 1 are larger than those of group 2. However, regional differences for LD regarding the height of peaks may be explained by the fact that in larger regions more patients are tested. Since the time to peak is based on comparisons over different time-points within regions the problem with variations between regions in population sizes is avoided.

4.3 Further comparison of the group of large cities with other regions

In section 4.1 it was found that the first LD case is reported approximately six weeks earlier in large cities than in the rest of the country, i.e. like for %ILI group 1 can be considered as a leading indicator for the incidence of the other group as well as for the whole of Sweden. In section 4.3.1 and 4.3.2 below a further description is made of the two groups and the whole of Sweden.

4.3.1 Parametric models of the expected incidence

The piecewise exponential curve presented in section 3.3.1 is used to estimate the curve of the expected value of LD. The estimated curves are shown in Figure 7.

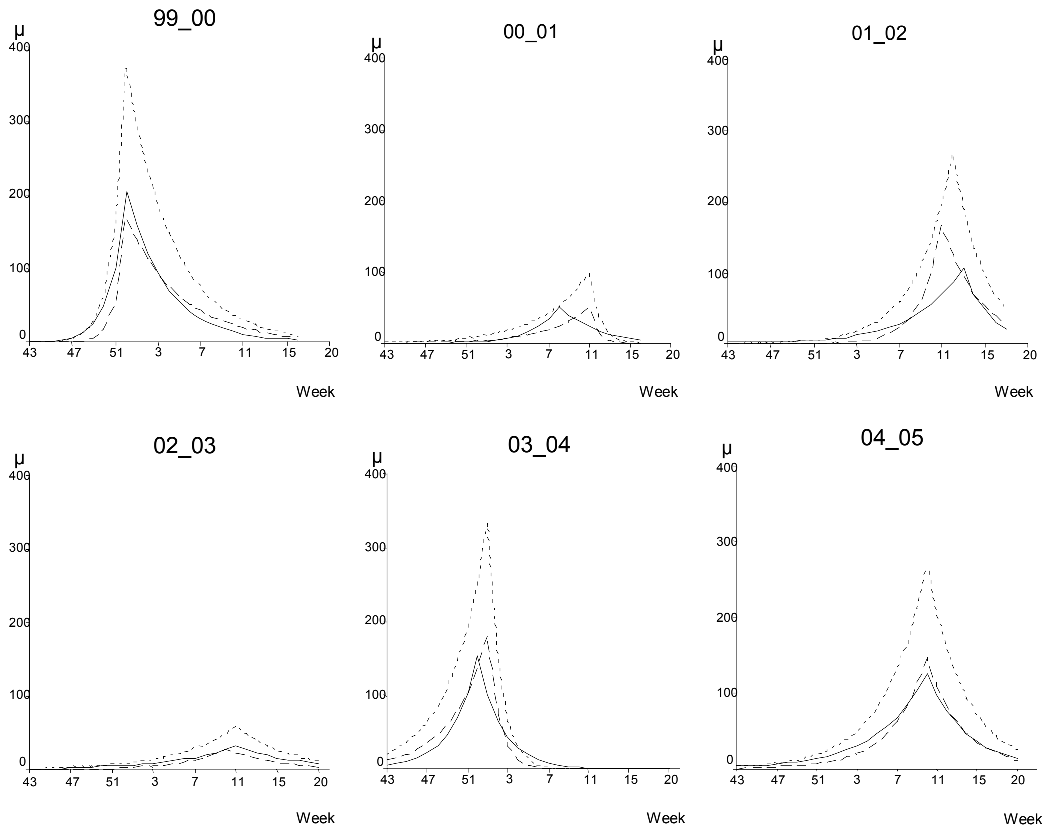


Figure 7. The exponential curves estimated for LD . Group 1 (——), group 2 (− − −), the whole of Sweden (- - - -).

It is hard to see that the group 1 has a tendency for earlier outbreak because of the scale of the vertical axis in Figure 7. Differences between group 1 and 2 regarding the time of the peak are seen.

4.3.2 Non-parametric estimates of the expected incidence

The estimated values of the incidence based on LD by the unimodal regression described in section 3.3.2 are given in Figure 8.

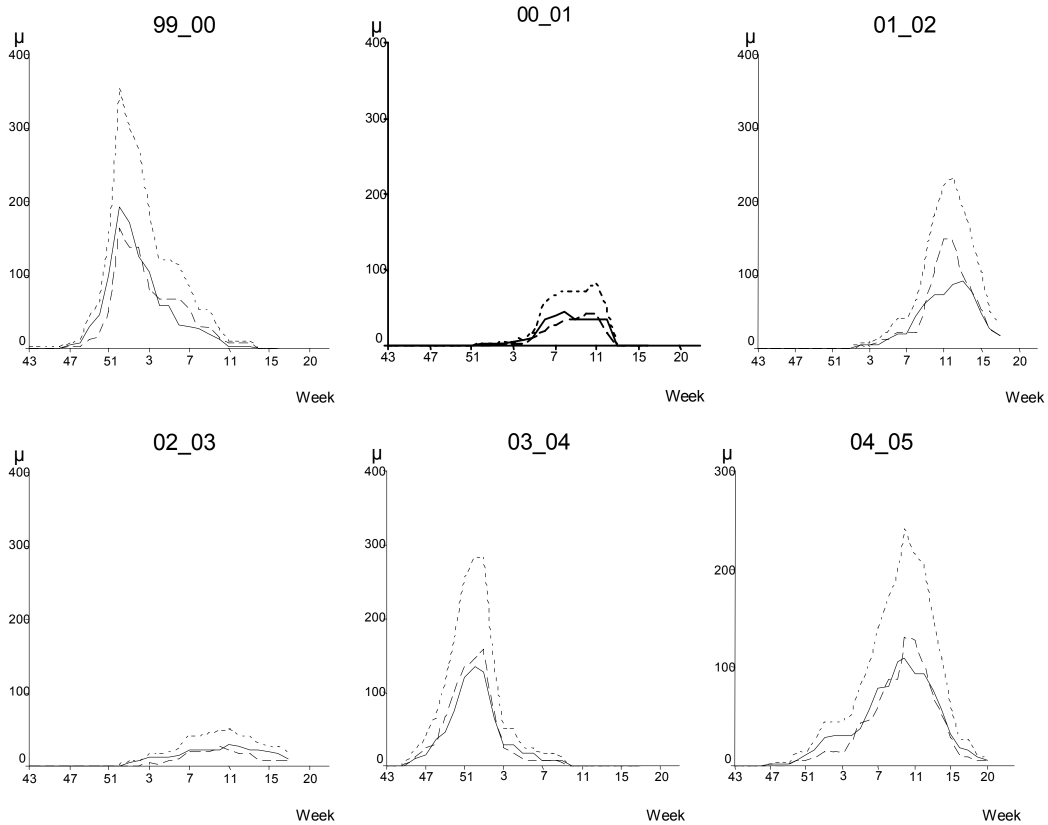


Figure 8. The estimated values of the incidence based on LD by the unimodal regression. Group 1 (——), group 2 (− − −), the whole of Sweden (- - - -).

5 CONCLUDING REMARKS

Spatial aspects of Swedish influenza data have been investigated. During the influenza season (autumn to early spring) data are collected weekly by the SMI. Two variables measuring the incidence of Swedish influenza have been analyzed, namely influenza-like-illness (ILI) and laboratory diagnoses (LD). Neither LD nor ILI is measured during the non-epidemic period (late spring to early autumn).

Information about the spatio-temporal patterns might give insight in the way the influenza spreads over Sweden. It might also be used in automated surveillance systems for outbreak and peak detection of the influenza. Several natural hypotheses about geographical patterns were examined but could not be verified as consistent over the years.

However, we found that for a group of large cities (Stockholm-Uppsala, Göteborg and Malmö), the outbreak of the influenza occurs at least four weeks earlier than for the rest of Sweden. The conclusion from the investigation in section 3.1 regarding the non-epidemic period and outbreak based on %ILI is that there is a time lag of about four weeks between the group of large cities, and the rest of Sweden. From the analysis in section 3.2 regarding the time of the peak of %ILI. the conclusion is that the patterns are rather weak.

The conclusion regarding the time of the first case of LD is that there is a difference between the groups. In the group of the large cities, the first case is reported about six weeks earlier than in the other group. The conclusion regarding the time of the peak of LD is that we did not find evidence for any pattern.

For further description of the two groups, parametric and non-parametric curves of the expected incidence were estimated for each of the groups and to aggregated data on %ILI and LD in section 3.3 and 4.3, respectively. The variances of the estimated incidence at each week will affect the variance of the alarm statistic used for surveillance and the variance of predictions. If one uses data from just one group the %ILI values will at each week be based on fewer regions than if one uses data for the whole of Sweden. If the differences between the two groups are small, then the variances of %ILI values will, for each of the two groups, be larger than the variances for the whole of Sweden. However, if the differences between the two groups are large, then the opposite may occur since much of the variability will be explained by the separation of the two groups. Accounting for the information of the differences between the two groups must be weighted against the loss of information by estimating the expected incidence with fewer regions. It is therefore an open question whether the leading property of the group of large cities may balance the smaller number of observations in this group when it comes to e.g. predictions and surveillance.

Construction of surveillance methods for spatial data is complicated and the spatial context leads to a multivariate surveillance situation. Spatial surveillance is beyond the scope of this paper and only a few strategies to construct alarm systems for outbreak detection are mentioned. For literature on spatial and multivariate surveillance, see e.g. Wessman (1998), Järpe (1999) and Sonesson and Frisén (2005). One aim might be to detect the outbreak in each group, and give separate alarms for different regions. Then we can perform separate and independent analyses for each of the groups. Another aim is to use spatial surveillance to answer the “overall” question of whether the Swedish influenza season has started yet. Some examples of how to construct a system for this latter question are given here:

1. Ignore the information on the time-lag and use aggregated data from all regions.

If there is no time-lag between the two regions then this is the best procedure. If there is a slight and variable time-lag, then further analyzes are required to examine the performance of aggregation. In the present data there seems to be a time-lag in the outbreak, which is relatively constant and rather large.

2. Use only data from the leading group in the surveillance. An alarm for outbreak in Sweden is based on an alarm for the leading group.

In this procedure the information on the time-lag is accounted for but other information is ignored by not using all the observations. Further analyzes are required to evaluate this method.

3. Monitor each of the groups separately. When there is an alarm in at least one of the groups we say there is an alarm for influenza in Sweden.

Then, information on the time-lag is accounted for and all the observations are used, though not necessarily in an optimal way. Further analyzes are required to evaluate this method.

4. Combine the data from the two groups to a univariate statistic (e.g. by a combination of time-lagged alarm statistics).

Further analyzes are required to optimize and evaluate this method.

This is an explorative analysis which might be the base for further research. The construction of methods using spatio-temporal information on the influenza spread, and the evaluation of such methods is a rather large and an interesting subject for further research.

ACKNOWLEDGEMENT

The data were made available to us by The Swedish Institute for Infectious Disease Control and we are grateful for discussions about the aims and the data quality. The authors are grateful for valuable comments by Professor Marianne Frisén and PhD. Eva Andersson. The research was supported by the Swedish Emergency Management Agency.

REFERENCES

Andersson, E., Bock, D. and Frisén, M. (2006) Exploratory analysis of Swedish influenza data. Research Report, Smittskyddsinstitutets rapportserie,

Andersson, Y., Normann, B. and Tideström, L. (1999) Fakta om smittsamma sjukdomar. Research Report, Swedish institute of infectious disease control, Stockholm

Bock, D., Andersson, E. and Frisén, M. (2006) On statistical surveillance of Swedish influenza incidence. Peak detection. Research Report, Smittskyddsinstitutets rapportserie, Fleming, D. M. and Cohen, J. M. (1996) Experience of European collaboration in influenza surveillance in the winter of 1993-1994. *Journal of Public Health Medicine,* **18,** 133-142.

Frisén, M. (1986) Unimodal regression. *The Statistician,* **35,** 479-485.

Ganestam, F., Lundborg, C. S., Grabowska, K., Cars, O. and Linde, A. (2003) Weekly antibiotic prescribing and influenza activity in Sweden: a study throughout five influenza seasons. *Scandinavian Journal of Infectious Diseases,* **35,** 836-842.

Järpe, E. (1999) Surveillance of the interaction parameter of the Ising model. *Communications in Statistics. Theory and Methods,* **28,** 3009-3027.

Järpe, E. (2000) On univariate and spatial surveillance. Ph.D Thesis. In *Department of Statistics* Göteborg University.

Järpe, E. (2001) Surveillance, environmental. In *Encyclopedia of Environmetrics*(Eds, El-Shaarawi, A. and Piegorsh, W. W.) Wiley, Chichester.

Linde, A., Brytting, M., Johansson, M., Wiman, Å., Högberg, L. and Ekdahl, K. (2004) Annual Report September 2003 - August 2004: The National Influenza Reference Center. Research Report, Swedish Institute for Infectious Disease Control, Stockholm

Linde, A. and Ekdahl, K. (2003) A/Fujian - föränderligt influensavirus som kan orsaka problem i vinter. *Läkartidningen,* **101,** 157.

Sonesson, C. and Frisén, M. (2005) Multivariate surveillance. In *Spatial surveillance for public health*(Eds, Lawson, A. and Kleinman, K.) Wiley, New York, pp. 169-186.

Wessman, P. (1998) Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics. Theory and Methods,* **27,** 1143-1161.

Research Report

2007:1	Andersson, E.:	Effect of dependency in systems for multivariate surveillance.
2007:2	Frisén, M.:	Optimal Sequential Surveillance for Finance, Public Health and other areas.
2007:3	Bock, D.:	Consequences of using the probability of a false alarm as the false alarm measure.
2007:4	Frisén, M.:	Principles for Multivariate Surveillance.
2007:5	Andersson, E., Bock, D. & Frisén, M.:	Modeling influenza incidence for the purpose of on-line monitoring.
2007:6	Bock, D., Andersson, E. & Frisén, M.:	Statistical Surveillance of Epidemics: Peak Detection of Influenza in Sweden.
2007:7	Andersson, E., Kühlmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S. & Frisén, M.:	Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden
2007:8	Bock, D., Andersson, E. & Frisén, M.:	Similarities and differences between statistical surveillance and certain decision rules in finance.
2007:9	Bock, D.:	Evaluations of likelihood based surveillance of volatility.