Modelling the spatial patterns of influenza incidence in Sweden

L. Schiöler

Modelling the spatial patterns

of influenza incidence in Sweden

Linus Schiöler

Statistical Research Unit, Department of Economics, University of Gothenburg,
SE 405 30 Göteborg, Sweden
E-mail: linus.schioler@statistics.gu.se

Abstract

Information about the spatial spread of epidemics can be useful for many purposes. The spatial aspect of Swedish influenza data was analyzed with the main aim of finding patterns that could be useful for statistical surveillance of the outbreak, i.e. for detecting an increase in incidence as soon as possible. In Sweden, two types of data are collected during the influenza season: laboratory diagnosed cases (LDI), collected by a number of laboratories, and cases of influenza-like illness (ILI), collected by a number of selected physicians. Quality problems were found for both types of data but were most severe for ILI. No evidence for a geographical pattern was found. Instead, it was found that the influenza outbreak starts at about the same time in the major cities and then occurs in the rest of the country. The data were divided into two groups, a metropolitan group representing the major cities and a locality group representing the rest of the country. The properties of the metropolitan group and the locality group were studied and it was found that the time difference in the onset of the outbreak was about one week. Both parametric and nonparametric regression models were suggested.

1. Introduction

Influenza is an epidemic disease which causes a significant number of deaths, especially among elderly people and infants, and also causes a considerable amount of absenteeism (see for example Szucs (1999) and Molinari *et al.* (2007)).

It is important to detect the onset of the outbreak as soon as possible, in order to be able to allocate the proper resources to the primary care sector and take preventive action. Statistical methods for surveillance increase the chances of early and correct detection. Automatic surveillance systems are now implemented in Sweden Cakici *et al.* (2010) and other countries. The three methods implemented in Sweden so far are based on Farrington *et al.* (1996, Frisén and Andersson (2009, Frisén *et al.* (2009, Kulldorff (1997). In Frisén and Andersson (2009, Frisén*, et al.* (2009) the application of one of the methods to influenza in Sweden is described. This method is applied to the country as a whole. Further development of the methods by incorporating spatial patterns may be beneficial for a surveillance system. There may be a time lag between the outbreaks in different regions of the country, and hence it may be possible to detect an outbreak earlier by considering spatial differences. At a regional level the number of reported influenza cases is small in Sweden, hence some aggregation of data is beneficial. A spatial pattern can be the base for such aggregation. The surveillance of spatial clusters of adverse health events has been analyzed for example by Kulldorff (2001) and Sonesson (2007). However, in Sweden data are available only for large regions which are not suitable for cluster analysis.

There are some earlier papers on influenza in Sweden. Bock and Pettersson (2006) also study the regional differences, but only up to the season 04/05. Their focus is on the peak and other techniques are used. Most papers concern the surveillance of the entire country. In Andersson et al. (2007) the problem of modelling influenza data is investigated. Bock et al. (2008) suggest a method for peak detection and apply it to Swedish data. Frisén and Andersson (2007) and Frisén et al. (2008) suggest a method for outbreak detection and apply it to Swedish influenza data. There is also some work on other related aspects of influenza in Sweden. Andersson *et al.* (2008b) propose a method for predicting the time and height of the peak of the influenza season. Ganestam et al. (2003) investigate the relation between influenza activity and the use of antibiotics. Uhnoo et al. (2003) describe the use of antiviral drugs and vaccines in the treatment and prevention of influenza. Grabowska et al. (2006) study the relation between influenza and Invasive Pneumococcal Disease. There are also

yearly and weekly influenza reports available from the Swedish Institute for Infectious Disease Control (SMI), at www.smittskyddsinstitutet.se.

In this report two different types of data on influenza are analyzed: cases verified in laboratories and cases of influenza-like illness (ILI) collected by the sentinel system. The laboratory diagnosed influenza (LDI) cases are identified at a number of laboratories: five virus laboratories at the university hospitals and SMI, and about 20 other microbiology laboratories. The number of laboratories participating varies from year to year. The sentinel system consists of about a hundred selected general practitioners who report the number of patients with influenza symptoms as well as the total number of visiting patients for each week. In order for a statistical surveillance system to be effective, it is important that the data collected are of sufficient quality, i.e. that they reflect the true state of the influenza incidence. The data are described and the potential quality problems of the data at hand are investigated in Section 2. Conclusions are drawn about the usefulness of the data for surveillance.

Spatial patterns for LDI are investigated in Section 4. The differences between the metropolitan areas and the rest of the country are reported in Section 5.

The modelling of the influenza incidence is important for effective statistical surveillance. Since the variation between years is large, a robust nonparametric or semiparametric model is suitable. A parametric model is needed for simulating data for evaluation purposes. In Section 6 we consider both parametric and semiparametric models.

In Section 7 some concluding remarks are made.

2. Data on influenza incidence

2.1. Influenza-like-illness

About a hundred selected physicians each week report the number of patients with influenza symptoms (#ILI) and the total number of visiting patients to SMI. The reporting of influenza starts at week 40 and is done on a voluntarily basis. Further information of the reporting can be found in Brytting *et al.* (2006b) and at the website of SMI (www.smittskyddsinstitutet.se).

SMI uses the percentage of the total number of visiting patients with ILI (%ILI) in the reporting. There is a large variation in the number of visiting patients, and as a consequence, %ILI may be somewhat unreliable as indicator of the influenza. Also, most regions have several weeks with missing values each year, both for the number of visiting patients and for the number of patients with influenza symptoms. It is not possible to tell whether the non-reporting units did not have any cases or if there are other reasons for the omission of the report. The inconstancies in the reporting, is most evident in the beginning and end of the season. This may be both because of the lack of cases and the physicians' expectation that there is no influenza present. Recently, there are laboratory analyses of samples of the ILI diagnosis. This might give more information of the usefulness of ILI data.

The low number of ILI cases and the variation in the number of visiting patients make surveillance at a regional level unfeasible. Furthermore, due to technical problems at SMI the number of patients of each region was unavailable for seasons after 04_05. Thus, meaningful aggregation of %ILI for different regions was not possible for later seasons. The ILI data could therefore not be used for spatial surveillance.

2.2. Web and telephone data

Due to the increased use of the internet it has been suggested that data over internet searches can be used as a proxy for the traditional types of data. Ginsberg *et al.* (2009) describes using Google's search data, and data is available for Sweden on www.google.org/flutrends/. Hulth *et al.* (2009) uses the search data from a website offering medical advice. The website is owned

by the Stockholm County Council and is aimed primarily at the residents of Stockholm. Neither of these sources offers spatial information.

The possibility of collecting data by telephone surveys and self-reporting has also been investigated by the SMI in Payne *et al.* (2005), Brytting *et al.* (2006a) and Bexelius *et al.* (2009), but are not collected on a regular basis.

2.3. *Laboratory diagnosed influenza cases*

2.3.1. Collection of data

The laboratory cases are reported by five viral laboratories and a number of microbiology laboratories. In general there is one laboratory in each larger city. In Stockholm there are two laboratories, one at Huddinge University Hospital (HS) and one at Karolinska University Hospital (KS). The number of reporting laboratories has increased but varies slightly between the years, as shown in Table I.

There are three different types of influenza viruses (A, B and C), which all belong to the group orthomyxoviridae. The typical influenza disease is mainly caused by influenza virus A and B, thus these are the types that will be studied. Most years there is a higher incidence for type A, and some years there are almost no cases of type B. There may be differences in the spread of A and B, for example the time of the peak differed slightly most years, but there was no consistent pattern in any direction. Because of the scarce data material we will use the sum of A and B in our analysis.

Table 1. Number of laboratories which has reported confirmed cases to SMI

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09
Number of laboratories	17	18	20	21	24	24	25	23	25	25

2.3.2. Quality problems

As with ILI the number of cases is relatively few, especially in the beginning and end of the season. A possible explanation is that there may be less inclination to perform laboratory testing if there is an expectation that the season hasn't started or is over.

Another potential problem is that there may be differences in policies regarding testing in different administrative areas. There may also be a stronger inclination to perform testing at hospitals with active research on influenza.

The differences in population size in the catchment areas of the laboratories may also be a problem; the number of cases is expected to be greater for laboratories serving big populations. Thus, you have to be careful with drawing conclusions regarding the incidence from the number of confirmed cases; a higher number of cases can be caused both by a higher incidence and a bigger population. Although it's claimed in Brytting *et al.* (2006c) that the laboratories are relatively evenly distributed with regards to population, there is still some variation.

The variation in the participation by laboratories could also be a problem. In general there is a trend that the number of participating laboratories is increasing. However, many laboratories have some years missing from the reporting. We were unable to determine the cause of this. One possible reason is administrative changes; the same population may be tested by different laboratories in different years. This is an example of a problem with what is referred to as metadata in Wallgren and Wallgren (2007). Proper documentation of why the number of laboratory differs from year to year would be helpful. There are also other examples of missing metadata.

2.3.3. Conclusions about the usefulness of LDI for spatial outbreak detection

The data are complete (for the period 99_00 to 08_09) for more than half of the regions, including the largest cities (Table 2). The varying number of laboratories could be a problem for at method that relies on a baseline to distinguish between the epidemic and non-epidemic phase. However, since it's primarily smaller laboratories that are inconsistent, the variation between seasons is a larger problem. Therefore a non- or semi-parametric approach would be more suitable.

As with ILI the number of LDI cases in each laboratory is in general too small to conduct surveillance for small changes in each region. However, by combining results from different parts of the country in an efficient way, inference regarding the outbreak in the whole country might be done more efficiently. Contrary to ILI, the LDI data is adequate for performing aggregation. However, care should be taken to that the groups might have different underlying population.

It is probable that the laboratories are more consistent than the sentinel physicians in their reporting. However, there may still be bias caused by the number of tests that are performed, e.g. the physicians may not test for influenza if they do not believe that the season has started. Another possible problem is that a hospital with a research interest in influenza may perform more extensive testing and therefore get a higher number of confirmed cases.

The conclusion is that LDI is more suitable than ILI for further analysis of the spatial spread of the influenza.

3.   Description of data for different regions

The total number of cases each year is shown in Table 2. Laboratories in larger cities tend to report more cases. A large variation between years as well as inconsistent reporting by some laboratories can be noted.

Table 2. Total number of laboratory diagnosed influenza cases. Laboratories with data for all years are shown in the top of the table. These are sorted by median. Laboratories with consistent reporting for the latest years are reported in the middle and laboratories with inconsistent reporting in the bottom.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09	Median
KS	350	143	215	111	249	282	110	120	247	247	231
Malmö	196	36	149	73	201	359	209	263	158	460	198,5
HS	293	109	178	95	189	252	121	155	185	180	179
Umeå	210	115	195	62	139	165	67	148	98	88	127
Skövde	102	52	140	39	107	184	34	88	15	98	93
Örebro	170	32	83	19	101	76	28	73	55	93	74,5
Göteborg	71	38	47	32	66	41	96	116	146	294	68,5
Falun	65	31	114	20	144	93	44	67	43	101	66
Uppsala	117	47	77	18	34	116	24	36	27	61	41,5
Halmstad	90	18	37	11	42	62	38	52	38	69	40
Karlstad	131	6	40	10	29	73	18	42	13	36	32,5
Kalmar	51	5	36	5	41	91	15	7	25	50	30,5
Linköping	31	5	32	24	23	17	9	16	14	24	20
Uddevalla	66	13	25	9	27	44	12	21	15	18	19,5
Västerås	10	1	9	2	28	29	10	26	4	13	10
Sundsvall		5	51	5	60	46	5	45	51	31	45
Gävle			5	4	15	14	14	20	11	16	14
Karlskrona			9	4	4	15	5	12	2	27	7
Eskilstuna				2	15	10	2	5	18	15	10
Borås					24	14	7	8	11	21	12,5
Jönköping					12	6	10	24	8	26	11
Kristianstad							7	27	16	54	21,5
Lund									26	61	43,5
Helsingborg									15	25	20
Luleå	22			2	15	14	16		5	6	14
Växjö	32	12	46	7	7	1	1				7
Östersund			9		1	15		1			5
Kungshamn		5									5
Trollhättan							2				2

Table 3 shows the number of weeks to the first laboratory diagnosed influenza case. There is considerable variation between the years and also between laboratories. One reason for the latter could be differences in population size. There may also be differences in incidence depending on population characteristics, such as the age distribution, as well as differences in testing policies. The largest cities, Stockholm, Göteborg and Malmö, have generally been

among the first to report cases. Umeå is also generally found among the cities with the earliest reports. Table 3 also shows the median number of weeks until the cumulative number of LDI cases exceeded 5.

Since the catchment areas of the laboratories differ, the reason that the larger cities reach a larger cumulative sum than the smaller cities could be either that the outbreak occurs earlier in the larger cities or that the probability of a large number is greater for a large population, or a combination of the two. This question will be further studied in Section 5.2.

Table 3. Number of weeks (since week 40) to the first laboratory diagnosed influenza case. The regions are sorted with respect to the median week for the first case. The median number of weeks until the cumulative number of LDI exceeds 5 is also shown as the last column.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09	Median	Median #>5
Göteborg	9	14	14	6	6	6	6	1	0	2	6	14.0
KS	3	14	7	8	5	7	11	10	4	0	7	12.0
HS	3	17	8	13	3	7	8	8	2	6	7.5	12.5
Umeå	3	17	15	12	7	10	5	3	8	7	7.5	14.0
Malmö	3	12	10	15	8	8	13	12	4	5	9	14.0
Borås					6	13	17	14	6	6	9.5	21.0
Skövde	8	14	15	4	5	13	16	8	14	8	10.5	16.5
Lund									11	11	11	15.0
Uppsala	4	14	14	15	8	3	18	11	7	11	11	15.5
Halmstad	9	18	14	17	7	9	14	16	2	2	11.5	18.5
Örebro	10	12	16	18	6	10	20	13	13	8	12.5	18.0
Helsingborg									14	12	13	16.0
Karlstad	6	19	14	15	8	11	17	12	17	3	13	17.0
Luleå	11			12	10	16	17		23	13	13	18.0
Falun	10	17	17	13	8	14	14	12	14	5	13.5	17.0
Jönköping					11	24	14	19	9	13	13.5	20.5
Kristianstad							15	12	18	6	13.5	20.5
Uddevalla	11	16	16	19	7	9	17	16	11	10	13.5	19.0
Sundsvall		21	14	20	8	16	16	11	13	11	14	17.5
Linköping	9	18	18	19	5	10	10	16	16	13	14.5	18.0
Eskilstuna			15	7	18	24	22	15	14	15	18.0
Västerås	12	23	22	20	9	11	18	8	19	3	15	17.0
Kalmar	9	20	16	23	5	16	14	19	15	13	15.5	19.0
Karlskrona			16	16	7	19	13	15	17	11	15.5	22.0
Gävle			18	17	3	16	17	18	10	10	16.5	19.0
Växjö	12	18	16	18	7	25	17				17	21.0
Östersund			19		20	14		18			18.5	23.5

4.   Spatial pattern

4.1.   Geographical position

Spatial analysis often concerns clusters. However, regional data on influenza in Sweden are available only for 25 large regions, which we found unsuitable for standard cluster analysis. Thus, we studied the possible spread to neighbouring areas by analyzing how the geographical position indicated by latitude and longitude is associated with the time of the outbreak. Table IV shows the correlations between the coordinates and the number of weeks until the number of LDI cases exceeded 5. None of these correlations differed significantly from zero.

Table 4. Spearman correlation between coordinates and number of weeks until LDI exceeded 5.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09	Median
Latitude	-,035	,177	-,126	-,261	,217	-,129	-,144	-,348	,007	,134	-,090
Longitude	-,021	-,046	-,290	-,431	,291	-,200	-,003	-,146	-,133	,149	-,177

Simultaneous analysis of geographical position and other variables will be reported in Section 4.3.

4.2.   Metropolitan and locality regions

In the tables above, we found that the large cities with good communications with other countries have a different pattern than the rest. We will examine classification into two groups: a metropolitan group consisting of Stockholm including Uppsala, Göteborg, Malmö and Umeå, and a locality group consisting of the rest of Sweden. Stockholm, Göteborg and Malmö all have considerably larger populations than the other cities, and they are part of the metropolitan areas as defined in Statistiska centralbyrån (2005). Uppsala, on the other hand, is more similar in population size to the cities in the locality group. However, the proximity and transport connections to Stockholm make Uppsala suitable to include in the metropolitan group. Moreover, the international airport of Arlanda is situated about halfway between Stockholm and Uppsala. We also included Umeå in the metropolitan group, although the city has a smaller population than the other cities in the group. Umeå is the largest city in the region of Norrland, which comprises about 59 % of the total area and 16% of the population of Sweden. The region's largest hospital is found here. Figure 1 shows the number of LDI cases for each group.

Using Spearman's rank correlation, we found that the pairwise correlations of weekly numbers of LDI cases in Stockholm, Göteborg, Malmö and Umeå were high (correlation coefficient >0.7 for most years). The correlation between Uppsala and the rest of the group was slightly lower but still high enough for it to be reasonable to include Uppsala in the group.

It could be argued that Lund and Borås should also be included in the metropolitan group, due to their proximity to Malmö and Göteborg, respectively. However, the reporting from Borås and Lund was inconsistent. There were also other quality problems associated with the reports from these cities. We chose to exclude them from the metropolitan group.
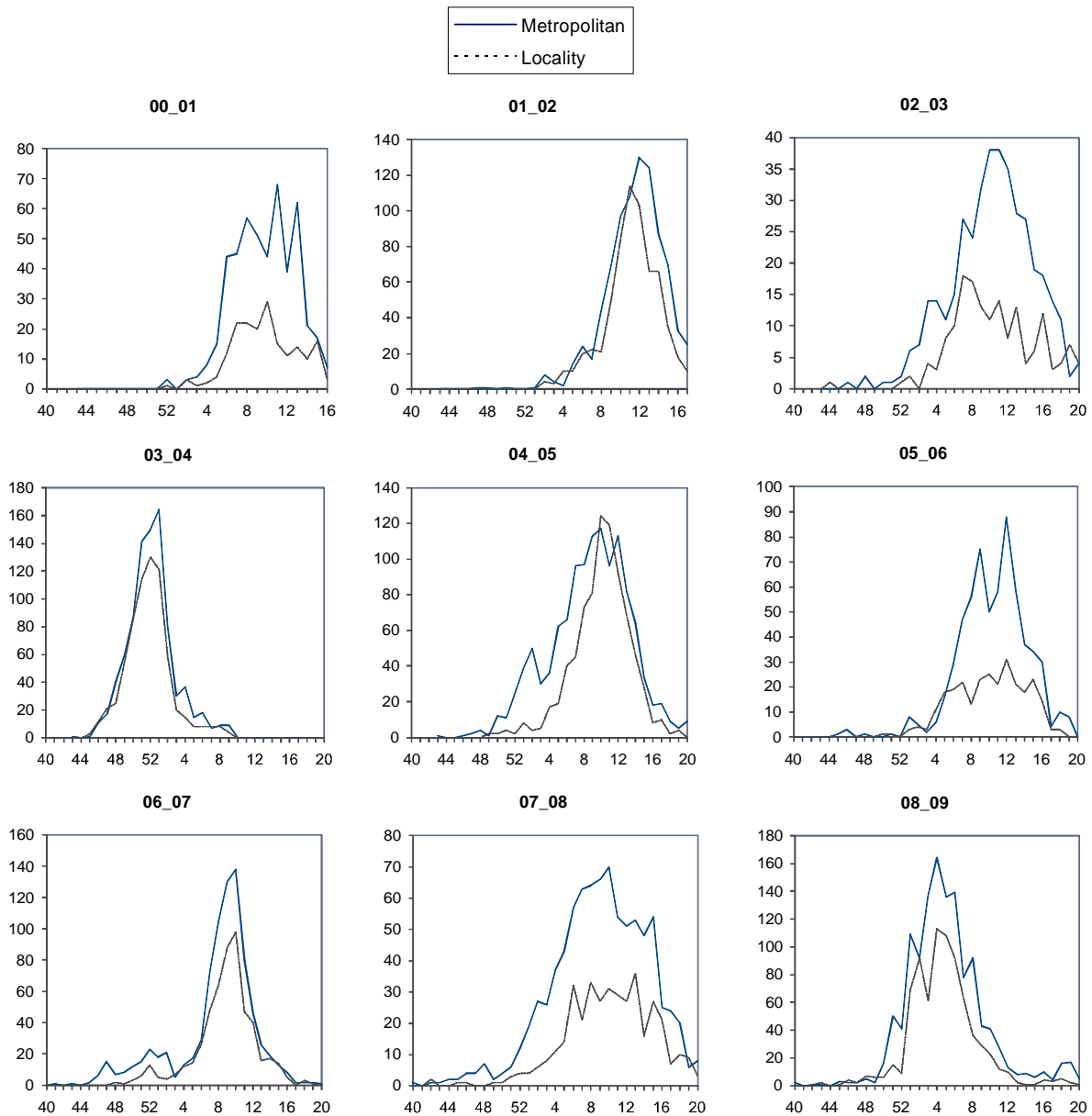
Figure 1. Number of laboratory diagnosed cases for the metropolitan group, Stockholm/ Uppsala, Göteborg, and Malmö (solid line) and the locality group, the rest of Sweden (dotted line).

4.3. Analysis of the spatial pattern

We aimed at finding which variables had the strongest influence on the time of the onset. To avoid interaction with missing data, only data from laboratories with data for all years were used. Different linear models with the time of the onset as dependent variable were analyzed. Year and group were used as qualitative factors and coordinates (latitude and longitude) as continuous variables. The results for one of the models are shown in table 5. We found that the group factor gave the highest partial coefficient of determination apart from year. The latitude and longitude coordinates were not significant in any of the models. Our conclusion was that there was no strong relation between the coordinates and the time of outbreak.

9

Table 5. Linear model with time of onset as dependent variable.

Source	DF	Type III SS	Mean Square	F Value	Pr > F	Partial R^2
Latitude	1	1.12	1.12	0.19	0.663	0.002
Longitude	1	3.34	3.34	0.57	0.453	0.005
Year	9	1669.24	185.47	31.53	<0.001	0.726
Group	1	94.74	94.74	16.11	<0.001	0.131

5.   Comparisons between the metropolitan areas and the rest of the country

5.1.   Differences in time of start of increase of incidence

In Table 6 the number of weeks until the cumulative number of LDI cases exceeded 10 is shown. This happened first in the metropolitan group in all years except 2002-2003.

Table 6. The number of weeks until the cumulative number of LDI cases exceeded 10.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09
Locality	17	16	16	6	13	15	11	12	8	9
Metropolitan	16	14	13	6	10	13	7	6	7	6
Difference	1	2	3	0	3	2	4	6	1	3

Table 6 suggests that there is a time lag between the two groups. Additional analyses were performed on each season to see which shift in time would make the incidence for the metropolitan, and locality areas more alike. The deviation between the groups were measured by the total root mean square deviation, RMSD=$\left[\frac{1}{n} \sum_{t=1}^{n} \left(M(t) - L(t+q) \right)^2 \right]^{1/2}$, where $M(t)$ and $L(t)$ denote the observation t of the metropolitan and locality group, respectively and q is the time lag. Hence a low value of RMSD is an indicator that the incidences in the two groups agree. Since our primary interest is the outbreak, we used only the observations from the start and until the number of observed cases in the metropolitan group had exceeded 15. We tried different time lags In the presence of a time lag we would expect the RMSD to be least for the correct value of q. All observations from the start until the metropolitan LDI exceeded 15 was used for lag zero. Later weeks was added to the locality group to get corresponding lagged values. The results are shown in Table 7. The RMSD calculated for all seasons was lowest for a lag of one week.

Table 7. Root mean square deviation between the metropolitan and locality groups.

Lag	RMSD
0	5.75
1	5.15
2	6.95
3	14.61

5.2. Size of groups

The uptake area of each laboratory is not known and therefore population size cannot be used in the analysis. A larger population means that a fixed number of cases will be exceeded earlier, even if the incidences are the same. The number of cases was larger for the metropolitan group. The median number of cases at the peak of the incidence was 123.5 for the metropolitan group and 105.5 for the locality group, a ratio of 1.17. To study the effect of the difference in size, we adjusted the size of the groups in the parametric model defined below and compared the time it took for the cumulative sum to exceed 5. The resulting time difference after the adjustment was about one day. Thus, a difference in population size of this magnitude could not be seen as the full explanation for the observed difference in the time of outbreak.

6. Models

Since the amount and quality of data is limited, and the variation between seasons is large constructing a parametric model of the influenza outbreak is hard. However, a simple model can be useful to study the properties of surveillance systems. In Andersson *et al.* (2008a) it was suggested that the Swedish influenza incidence could be modelled by a Poisson process with the intensity following an exponential curve.

6.1. Parametric models of the expected incidence

A parametric model is useful to describe details of the outbreak. In order to make a simulation study of the properties of a surveillance method, some sort of parametric model is also needed. In Frisén and Andersson (2009) the model

$$\mu(t) = \begin{cases} \mu_0, & t < \tau \\ \exp(\beta_0 + \beta_1(t - \tau + 1)), & t \geq \tau \end{cases},$$

where τ denotes the time of the onset, is used for a typical curve of the total number of LDI cases in the whole of Sweden. The constant phase, μ_0, was roughly estimated to $\mu_0 = 1$ from Swedish LDI data for eight years. The model was estimated from the incidence in the season 03/04, when the outbreak was neither particularly severe nor particularly mild. The estimates of the parameters were $\beta_0 = -0.26$ and $\beta_1 = 0.826$.

By the results above we have that the locality and metropolitan groups each had about half the number of cases in Sweden as a whole and an approximate time lag between them of about one week. Thus, the relation between the incidences of the total (T), metropolitan (M) and locality (L) areas can be expressed by

$$\mu_T(t) = \mu_M(t) + \mu_L(t) = \begin{cases} \mu_0, & t < \tau_M \\ \exp(\beta_0^* + \beta_1^*(t - \tau_M + 1)) + \mu_0/2, & \tau_M \leq t < \tau_L \\ \exp(\beta_0^* + \beta_1^*(t - \tau_M + 1)) + \exp(\beta_0^* + \beta_1^*(t - \tau_L + 1)) & t \geq \tau_L \end{cases}$$

where $\tau_L = \tau_M + 1$ and $\mu_0 = 1$. The parameters $\beta_0^* = -0.62$ and $\beta_1^* = 0.826$ give a good approximation of the model for the total incidence above. This curve fitted well to the data for the same season (03/04) for some values of the starting time. It also fitted rather well for some other seasons, while a good fit for all seasons could not be expected due to the marked differences between the seasons.

6.2. Nonparametric models of the expected incidence

Due to the limited quality and the variation between years, the parametric model is unsuitable for inference. The interaction between the estimates of the start and slope of the outbreak is

another weakness of parametric models. The use of order restrictions for modelling outbreaks is suggested in Frisén *et al.* (2010), where it is assumed that the incidence is constant up to some starting point and then non-decreasing. A similar assumption is used in Andersson, *et al.* (2008b), where the time of onset and the slope are used for predicting the time and height of the peak in influenza incidence. The time difference between the (interpolated) time points when the total number of LDI cases in Sweden exceeds 30 and 10, respectively, is used as an indicator of the slope. We applied these techniques to the aggregated data but used the time difference between 15 and 5, since each of the groups accounts for about half of the total number of cases in Sweden. We found no significant difference between the slopes of the metropolitan and locality groups.

6.3.   Semiparametric model

The nonparametric model by order restriction can be combined with the Poisson distribution to a semiparametric model. In Frisén, *et al.* (2009) a semiparametric method of surveillance is applied to Swedish LDI data for the country as a whole. Figure 2 shows the alarm statistic of the method applied to the metropolitan and locality groups. The metropolitan group had a tendency to an earlier increase than the locality group. Thus, an earlier alarm or first warning can be expected here.
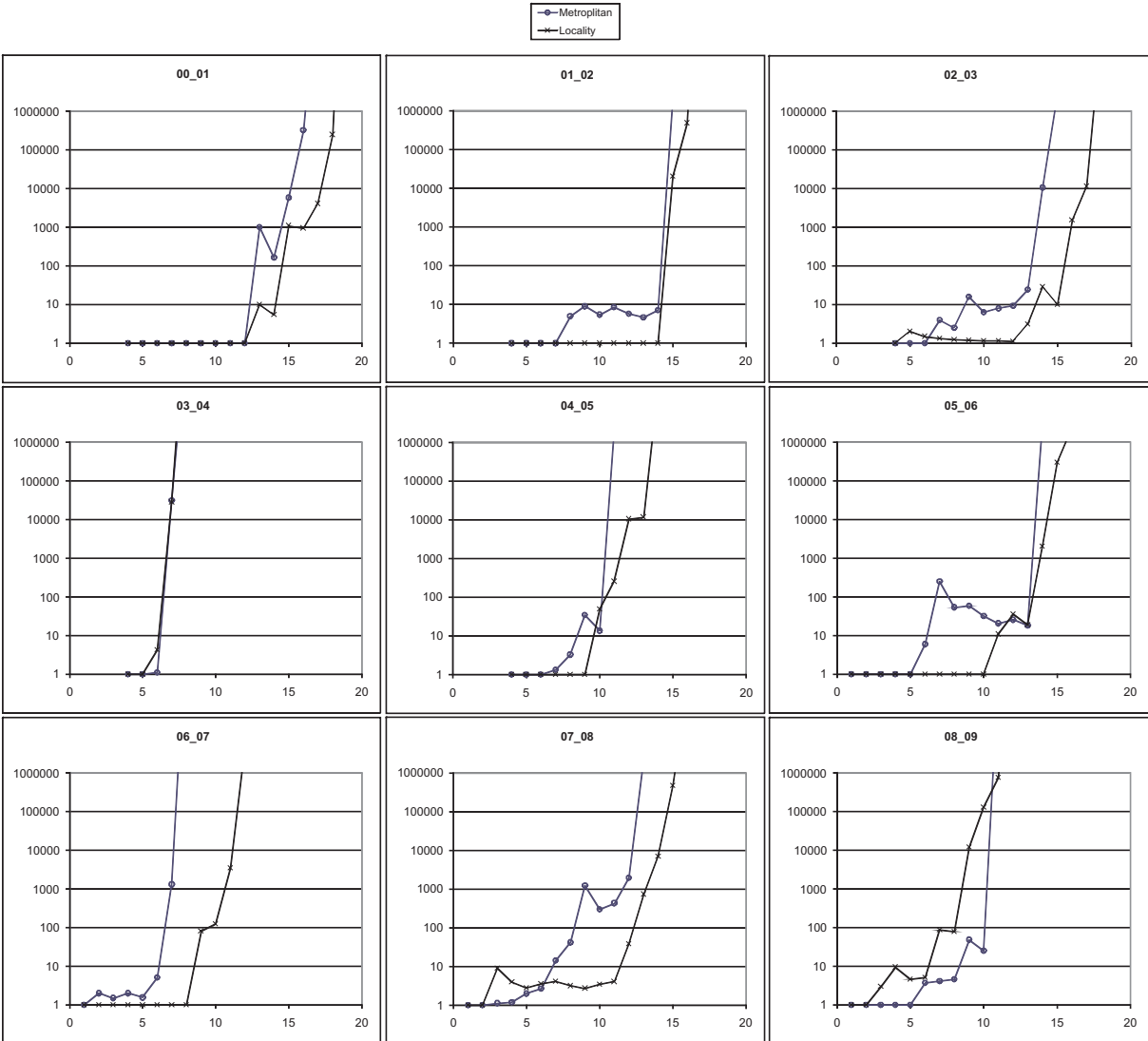

Figure 2. OutP alarm statistics for the groups. The dots represents the metropolitan group and the crosses represent the locality group.

12

7.   Concluding remarks

The surveillance of infectious diseases such as influenza has drawn much attention recently. We analyzed the spatial aspect of Swedish influenza data with the main aim of finding patterns that could be useful for statistical surveillance of the outbreak, i.e. for detecting an increase in incidence as soon as possible.

In Sweden, several types of data are collected during the influenza season. The most established ones are data on laboratory diagnosed cases (LDI), collected by a number of laboratories, and cases of influenza-like illness (ILI), collected by a number of selected physicians. Quality problems were found for both types of data but were most severe for ILI. A potential problem with LDI data is that policies regarding testing may differ between administrative areas. Hospitals conducting research on influenza may also be more inclined to perform testing. The differences in population size between the catchment areas of the laboratories may also constitute a problem. Since the population size was not known we could not adjust for this. The number of cases can be expected to be greater for laboratories serving large populations. Thus, one has to be careful with drawing conclusions regarding the incidence from the number of confirmed cases, since a higher number of cases can be the result of both a higher incidence and a larger population. The varying number of reporting laboratories may also be a problem, particularly when using a surveillance method that relies on a baseline to distinguish between the epidemic and non-epidemic phases. However, the fact that primarily smaller laboratories are inconsistent in their reporting lessens this effect.

In Frisén and Andersson (2009) and Frisén, *et al.* (2009) it has been shown that Swedish influenza data can be useful for surveillance. By combining results from different parts of the country in an efficient way, inference regarding the outbreak in the country as a whole might be performed more efficiently. We found that there was a time lag between the metropolitan and locality areas. This can be potentially useful for faster and more reliable detection of the outbreak.

Spatial patterns such as those based on geographical coordinates were examined. We found no evidence for a relation between the time of the onset of the outbreak and a location to the north/south or east/west. We found that in the major cities, Stockholm (including Uppsala), Göteborg, Malmö and Umeå, the onset of the influenza outbreak seemed to occur earlier than in the rest of the country. Analysis with respect to the variables coordinates, group (metropolitan/locality) and year revealed that year and group was the most important as concerns the time of the onset of the outbreak. These metropolitan regions all have major airports nearby, and commuting is common.

The properties of the metropolitan and the locality groups were analysed by studying the time at which a certain incidence was reached, the similarity between lagged variables, and graphs of the incidence and alarm statistic at the onset. Although the variation between years was quite large, a difference of one week between the metropolitan and locality groups was a good approximation for most years. There are a number of factors that could contribute to the difference in influenza incidence between regions. Lowen *et al.* (2007) found that temperature and humidity had an effect on the transmission of influenza virus. This may be a factor in Sweden due to its diverse climate. However, we found no influence of the geographical coordinates, which are of course correlated with climate variables. Brownstein et al. (2006) found that air travel had a significant effect on the spread of influenza in the USA. It is thus probable that major cities with well-developed means of transport may have an earlier outbreak than smaller cities.

Stochastic models for influenza incidence are needed for many purposes. Andersson, *et al.* (2008a) found that the Poisson distribution fits well to data at the onset of the outbreak. In this paper, parametric exponential regression models were suggested both for the country as a whole and for the metropolitan and locality groups separately. As for the incidence slope at

the onset, no evidence was found for a difference between the two groups. These parametric models are useful to generate data for simulation and for enhancing understanding. The variation in incidence between the years is large. Therefore, a nonparametric or semiparametric approach would be more suitable. For surveillance purposes, we suggest using a robust nonparametric regression model with order restriction.

Acknowledgements

References

Andersson, E., Bock, D. and Frisén, M. (2007) Modeling influenza incidence for the purpose of on-line monitoring. *Stat. Methods Med. Res.*

Andersson, E., Bock, D. and Frisén, M. (2008a) Modeling influenza incidence for the purpose of on-line monitoring. *Stat. Methods Med. Res.,* 421-438.

Andersson, E., Kuhlmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S. and Frisén, M. (2008b) Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden. *Scand. J. Public Health,* 475-482.

Bexelius, C., Merk, H., Sandin, S., Ekman, A., Nyrén, O., Kühlmann-Berenzon, S., Linde, A. and Litton, J.-E. (2009) SMS versus telephone interviews for epidemiological data collection: feasibility study estimating influenza vaccination coverage in the Swedish population. *Eur. J. Epidemiol.,* 73-81.

Bock, D., Andersson, E. and Frisén, M. (2008) Statistical surveillance of epidemics: Peak detection of influenza in Sweden. *Biometrical Journal,* 71-85.

Bock, D. and Pettersson, K. (2006) Exploratory analysis of spatial aspects on the Swedish influenza data. Smittskyddsinstitutets rapportserie 3:2006. Swedish Institute for Infectious Disease Control.

Brownstein, J. S., Wolfe, C. J. and Mandl, K. D. (2006) Empirical Evidence for the Effect of Airline Travel on Inter-Regional Influenza Spread in the United States. *PLoS Medicine,* e401.

Brytting, M., Stivers, M., Dahl, H., Serifler, F., Linde, A. and Rubinova, S. (2006a) Annual Report July 2006 - June 2007: The National Influenza Reference Center. Swedish Institute for Infectious Disease Control.

Brytting, M., Stivers, M., Dahl, H., Serifler, F., Linde, A. and Rubinova, S. (2006b) Annual Report july 2007- june 2008 : The National Influenza Reference Center. Swedish Institute for Infectious Disease Control.

Brytting, M., Stivers, M., Linde, A. and Rubinova, S. (2006c) Annual Report September 2005 - August 2006: The National Influenza Reference Center. Swedish Institute for Infectious Disease Control.

Cakici, B., Hebing, K., Grünewald, M., Saretok, P. and Hulth, A. (2010) CASE –a framework for computer supported outbreak detection. . *BMC Med Inform Decis Mak.*

Farrington, C. P., Andrews, N. J., Beal, A. D. and Catchpole, M. A. (1996) A statistical algorithm for the early detection of outbreaks of infectious disease. *J. R. Statist. Soc. A,* 547-563.

Frisén, M. and Andersson, E. (2007) Semiparametric surveillance of outbreaks. Research report 2007:11. Statistical Research Unit, Department of Economics, Göteborg University, Sweden.

Frisén, M. and Andersson, E. (2009) Semiparametric surveillance of monotonic changes. *Sequential Analysis,* 434-454.

Frisén, M., Andersson, E. and Pettersson, K. (2010) Semiparametric estimation of outbreak regression. *Statistics,* 107-117.

Frisén, M., Andersson, E. and Schiöler, L. (2008) Robust outbreak surveillance of epidemics in Sweden. *Statistics in Medicine,* in press.

Frisén, M., Andersson, E. and Schiöler, L. (2009) Robust outbreak surveillance of epidemics in Sweden. *Stat. Med.,* 476-493.

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,* 836-842.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009) Detecting influenza epidemics using search engine query data. *Nature,* 1012-1014.

Grabowska, K., Hogberg, L., Penttinen, P., Svensson, A. and Ekdahl, K. (2006) Occurrence of invasive pneumococcal disease and number of excess cases due to influenza. *BMC Infectious Diseases,* 58.

Hulth, A., Rydevik, G. and Linde, A. (2009) Web Queries as a Source for Syndromic Surveillance. *PLoS ONE,* e4378.

Kulldorff, M. (1997) A spatial scan statistic. *Comm. Stat. Theor. Meth.,* 1481-1496.

Kulldorff, M. (2001) Prospective time periodic geographical disease surveillance using a scan statistic. *J. R. Statist. Soc. A,* 61-72.

Lowen, A. C., Mubareka, S., Steel, J. and Palese, P. (2007) Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. *PLoS Pathogens,* e151.

Molinari, N.-A. M., Ortega-Sanchez, I. R., Messonnier, M. L., Thompson, W. W., Wortley, P. M., Weintraub, E. and Bridges, C. B. (2007) The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine,* 5086-5096.

Payne, L., Kühlmann-Berenzon, S., Ekdahl, K., Giesecke, J., Högberg, L. and Penttinen, P. (2005) 'Did you have flu last week?' A telephone survey to estimate a point prevalence of influenza in the Swedish population. *Eurosurveillance,* 241-244.

Sonesson, C. (2007) A CUSUM framework for detection of space-time disease clusters using scan statistics. *Stat. Med.,* 4770-4789.

Statistiska centralbyrån (2005) Geografin i statistiken - regionala indelningar i Sverige MIS 2005:2. Statistiska centralbyrån.

Szucs, T. (1999) The socio-economic burden of influenza. *Journal of Antimicrobial Chemotherapy,* 11-15.

Uhnoo, I., Linde, A., Pauksens, K., Lindberg, A., Eriksson, M. and Norrby, R. (2003) Treatment and prevention of influenza: Swedish recommendations. *Scandinavian Journal of Infectious Diseases,* 3-12.

Wallgren, A. and Wallgren, B. (2007) *Register-based statistics*, Chichester: Wiley.

Research Report

2007:10	Bock, D. & Pettersson, K.	Explorative analysis of spatial aspects on the Swedish influenza data.
2007:11	Frisén, M. & Andersson, E.	Semiparametric surveillance of outbreaks.
2007:12	Frisén, M., Andersson, E. & Schiöler, L.	Robust outbreak surveillance of epidemics in Sweden.
2007:13	Frisén, M., Andersson, E. & Pettersson, K.	Semiparametric estimation of outbreak regression.
2007:14	Pettersson, K.	Unimodal regression in the two-parameter exponential family with constant or known dispersion parameter.
2007:15	Pettersson, K.	On curve estimation under order restrictions.
2008:1	Frisén, M.	Introduction to financial surveillance.
2008:2	Jonsson, R.	When does Heckman's two-step procedure for censored data work and when does it not?
2008:3	Andersson, E.	Hotelling´s T2 Method in Multivariate On-Line Surveillance. On the Delay of an Alarm.
2008:4	Schiöler, L. & Frisén, M.	On statistical surveillance of the performance of fund managers.
2008:5	Schiöler, L.	Explorative analysis of spatial patterns of influenza incidences in Sweden 1999—2008.
2008:6	Schiöler, L.	Aspects of Surveillance of Outbreaks.
2008:7	Andersson, E & Frisén, M.	Statistiska varningssystem för hälsorisker
2009:1	Frisén, M., Andersson, E. & Schiöler, L.	Evaluation of Multivariate Surveillance
2009:2	Frisén, M., Andersson, E. & Schiöler, L.	Sufficient reduction in multivariate surveillance