Research Report
Statistical Research Unit
Department of Economics
University of Gothenburg
Sweden

Explorative analysis of spatial patterns of influenza incidences in Sweden 1999—2008

L. Schiöler

Research Report 2008:5
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			

Explorative analysis of spatial patterns of influenza incidences in Sweden 1999—2008

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

Summary

Information about the spatial spread of epidemics can be useful for many purposes. In this paper, the spatial aspect of Swedish influenza data is 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 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 two weeks. This justifies a different spatial model than the one usually used for infectious diseases.

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)). 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. An early detection could also be useful for taking preventive action. Here statistical surveillance is a valuable tool, as it increases the chances of an early and correct detection. The aim of this paper is primarily to examine spatial patterns that could be useful for a surveillance system.

In order for a surveillance system to be as effective as possible, it is important to consider spatiotemporal variations of the influenza epidemic. There may be a considerable time lag between different regions of the country, and hence it may be possible to detect an outbreak earlier by taking spatial differences into account. In Sweden the number of reported influenza cases is quite small. It would thus be useful to find some spatial pattern which could lead to a sufficient aggregation of data.

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, Bock, and Frisén (2007) the problem of modeling influenza data is investigated. Bock, Andersson, and Frisén (2008) suggest a method for peak detection and apply it to Swedish data. Frisén and Andersson (2007) and Frisén, Andersson, and Schiöler (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. (2008) 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 for ILI and Section 3 for LDI. Conclusions are drawn about the usefulness of the data for surveillance.

There are a number of factors that could contribute to the difference 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. Brownstein, Wolfe, and Mandl (2006) found that air travel had a significant effect on the spread of influenza in the USA. It is thus possible that major cities with well-developed means of communication may have an earlier outbreak than smaller cities. The surveillance of spatial clusters of adverse health events has been analyzed for example by Kulldorff (2001) and Sonesson (2007). Simple spatial patterns for LDI are discussed in Section 3.3. The differences between the metropolitan areas and the rest of the country are reported in Section 4.

In Section 5 some concluding remarks are made.

2. Influenza-like-illness

2.1. Collection of data

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. This reporting system is referred to as the sentinel system. Since it is not mandatory to report influenza in Sweden, the reporting is done on a voluntarily basis. The official reporting of influenza starts at week 40. Further information of the reporting can be found in Ganestam *et al.* (2003) and in the annual reports from The National Influenza Reference at the website of SMI (www.smittskyddsinstitutet.se).

2.2. Quality problems

SMI uses the percentage of the total number of visiting patients with ILI (%ILI) in the reporting. As can be seen in Figure 1 the variation in the number of visiting patients is large. As a consequence, %ILI may be somewhat unreliable as indicator of the influenza.

In the data available, most regions each year have several weeks with missing values, 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.

Due to these inconstancies in the reporting there is a high degree of error in %ILI for the individual regions. Hence, %ILI is not useful at a regional level.

The problem is most evident in the beginning and end of the season. This may be because of the lack of cases or the physicians' expectation that there is no influenza present. A consequence of this is that it's hard to estimate a reliable baseline for the non-epidemic period.

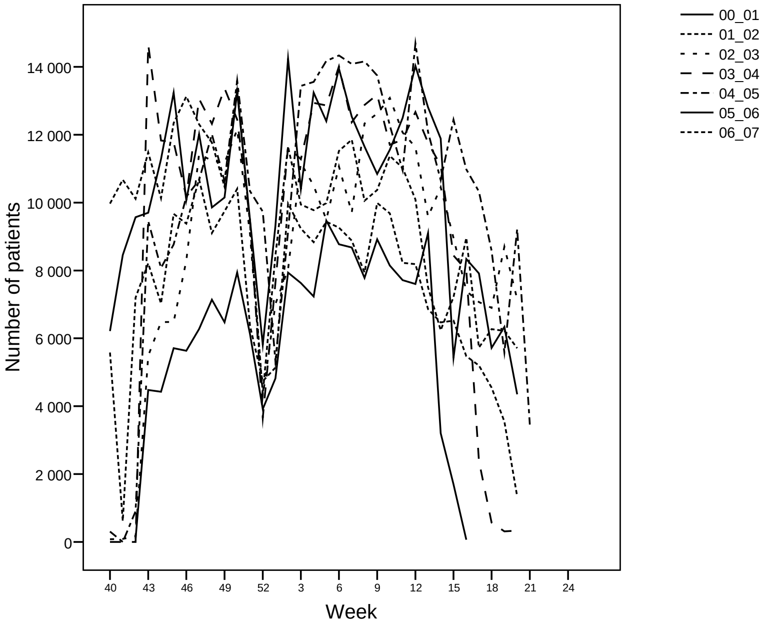

Figure 1. *The total number of patients reported by the sentinel physicians.*

2.3. Description of data

The percentage of ILI cases for the entire Sweden is shown in Figure 2. In Table 1 the number of weeks to the first reported case of influenza-like illness is shown. There is considerably variation between the regions. Due to the low incidence in the beginning of the season and the reporting bias mentioned above this could be expected.

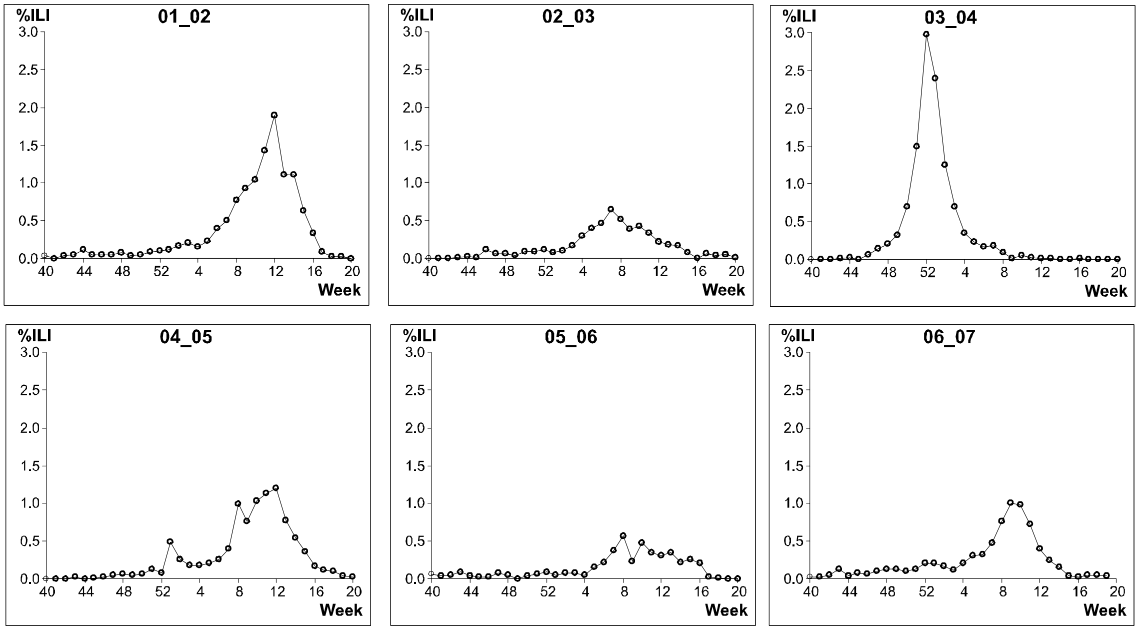

Figure. 2 *%ILI for the entire of Sweden.*

Table 1. Number of weeks (since week 40) to the first reported case of influenza-like illness. The regions are sorted with respect to the median.

Region	00_01	01_02	02_03	03_04	04_05	05_06	06_07	Median	Range
Dalarna	3	0	11	8	5	0	3	3	11
Stockholm	7	2	6	4	8	3	0	4	8
Skåne	5	4	6	9	8	0	3	5	9
Västerbotten	6	3	4	6	8	2	5	5	6
Västra Götaland	11	4	6	6	3	7	3	6	8
Blekinge	18	6	15	7	6	3	17	7	15
Uppsala	8	0	5	8	7	7	2	7	8
Värmland	7	4	3	4	12	14	13	7	11
Västmanland	3	8	6	8	11	0	8	8	11
Gävleborg	13	9	18	7	16	4	3	9	15
Jämtland		5	22	10	9	22	7	9.5	17
Kalmar	3	5	15	10	13	4	10	10	12
Södermanland	19	2	16	9	10	10	8	10	17
Jönköping	10	10	19	19	23	1	12	12	22
Västernorrland	15	16	10	3	13	19	13	13	16
Östergötland	17	23	16	13	8	14	3	14	20
Kronoberg	10	14	15	6	15	19	16	15	13
Halland	12	23	16	7	18	17	14	16	16
Norrbotten	18	18	10	7	13	18	32	18	25
Örebro	15	22	18	10	18		19	18	12

Table 2. Number of weeks (since week 40) until the maximum value of %ILI. The regions are sorted with respect to the median

Region	00_01	01_02	02_03	03_04	04_05	05_06	06_07	Median	Range
Västerbotten	17	23	19	15	14	24	13	17	11
Norrbotten	23	25	16	14	14	19	33	19	19
Stockholm	22	25	18	13	14	24	20	20	12
Gävle	22	22	20	15	25	21	12	21	13
Södermanland	23	28	20	14	16	21	21	21	14
Västra Götaland	20	25	20	13	24	21	22	21	12
Örebro	23	25	22	16	19		21	21.5	9
Blekinge	21	22	20	14	25	27	22	22	13
Dalarna	22	25	22	14	26	14	22	22	12
Halland	23	26	19	14	25	22	21	22	12
Uppsala	22	26	19	13	17	24	33	22	20
Västernorrland	19	22	24	10	26	26	22	22	16
Västmanland	22	24	20	13	30	24	22	22	17
Östergötland	24	27	22	14	25	18	21	22	13
Jämtland		25	24	12	23	23	23	23	13
Kronoberg	23	25	21	15	27	21	23	23	12
Värmland	23	23	24	13	24	21	21	23	11
Jönköping	24	24	22	20	24	23	28	24	8
Kalmar	25	24	24	13	25	21	25	24	12
Skåne	21	26	26	14	26	26	23	26	12

In Table 2 the number of weeks to the peak is shown. Some of the regions have a very low number of ILI cases, making the time of peak somewhat arbitrary for these regions. The height of the peak of %ILI is shown in Table 3. There is a considerable variation between the regions. Due to the variation in the reported number of visiting patients the value of %ILI is unreliable. The problem of low number of patients is present in this table as well. As an example, Norrbotten 02_03 has a peak of 100 percent, which in fact corresponds to only one patient with symptoms of ILI. This is probably caused by both incorrect and missing reports of the number of visiting patients from the sentinel physicians in Norrbotten that week. This illustrates why %ILI is an unreliable measure of incidence, especially at a regional level. In order to make it a useful measure some improvement in the reporting is needed.

2.4. Conclusions about the usefulness of ILI for outbreak detection

As mentioned above there are quality problems in the 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.

Table 3. The height of the peak of %ILI measured by the maximum value of %ILI. The regions are sorted with respect to the median.

Region	00_01	01_02	02_03	03_04	04_05	05_06	06_07	Median
Norrbotten	7.19	28.00	100.00	6.58	1.89	2.50	10.00	7.19
Blekinge	2.94	6.25	5.26	16.90	13.04	5.71	2.56	5.71
Västernorrland	4.17	5.06	6.25	6.78	8.89	2.82	1.78	5.06
Dalarna	19.35	4.66	1.08	10.29	6.77	1.65	2.07	4.66
Skåne	11.11	4.32	2.63	11.20	4.68	1.96	3.70	4.32
Jönköping	4.02	10.53	3.85	3.85	7.14	1.74	4.17	4.02
Stockholm	3.59	3.23	5.13	7.01	1.38	1.42	2.40	3.23
Västerbotten	4.48	9.73	1.65	3.18	3.08	1.69	4.17	3.18
Kronoberg	3.49	3.02	1.52	1.25	3.45	6.25	2.13	3.02
Södermanland	3.26	2.94	2.78	6.90	1.77	1.69	6.67	2.94
Värmland	3.55	2.77	0.42	1.42	4.50	2.04	5.56	2.77
Halland	3.26	1.54	2.68	11.71	2.40	1.20	0.94	2.40
Uppsala	3.83	2.35	1.45	5.30	1.50	0.30	2.61	2.35
Jämtland		1.90	1.99	2.30	0.77	1.03	2.88	1.95
Västra Götaland	3.55	1.80	0.39	1.95	1.99	0.65	0.75	1.80
Västmanland	3.12	1.69	0.29	3.82	1.40	0.28	1.12	1.40
Gävleborg	3.26	1.26	0.43	1.19	1.40	1.54	0.84	1.26
Örebro	1.74	1.24	1.83	1.40	0.23	0.00	0.65	1.24
Östergötland	1.43	0.31	0.29	1.50	1.21	0.41	0.53	0.53
Kalmar	1.20	3.45	0.40	0.45	1.13	0.23	0.27	0.45

3. Laboratory diagnosed influenza cases

3.1. Collection of data

The laboratory cases are reported from 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 varies slightly between the seasons, as shown in Table 4.

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 4. 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
Number of laboratories	17	16	20	21	24	24	25	23	25

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. (2006) 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.

3.3. Description of data

The total number of cases for each year is shown in Table 5. Larger cities tend to have more cases; Stockholm (laboratories HS and KS) has most cases every year. A noticeably exceptions is Umeå, which for many year has the second most cases. The number of cases for the entire Sweden is shown in Figure 3. A large variation between years can be seen.

Table 5. Total number of laboratory diagnosed influenza cases sorted by median. Laboratories with data for all years are shown in the top of the table, laboratories with consistent reporting for the latest years 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**	**Median**
KS	348	143	215	111	249	282	110	120	247	215
Malmö	196	36	149	73	201	359	209	263	158	196
HS	292	109	178	95	189	252	121	155	185	178
Umeå	213	115	195	62	139	165	67	148	98	139
Skövde	102	52	140	39	107	184	34	88	15	88
Örebro	169	32	83	19	101	76	28	73	55	73
Falun	67	31	114	20	144	93	44	67	43	67
Göteborg	86	38	47	32	66	41	96	116	146	66
Halmstad	75	18	37	11	42	62	38	52	38	38
Uppsala	116	47	77	18	34	116	24	36	27	36
Karlstad	131	6	40	10	29	73	18	42	13	29
Kalmar	51	5	36	5	41	91	15	7	25	25
Uddevalla	66	13	25	9	27	44	12	21	15	21
Linköping	32	5	32	24	23	17	9	16	14	17
Västerås	9	1	9	2	28	29	10	26	4	9
Sundsvall			51	5	60	46	5	45	51	46
Gävle			5	4	15	14	14	20	11	14
Karlskrona			9	4	4	15	5	12	2	5
Eskilstuna				2	15	10	2	5	18	7.5
Jönköping					12	6	10	24	8	10
Kristianstad							7	27	16	16
Lund									26	26
Helsingborg									15	15
Luleå	22			2	15	14	16		5	14.5
Borås					24	14	6		11	12.5
Växjö	32	12	46	7	7	1	1			7
Östersund			9		1	15		1		5
Trollhättan							3	8		5.5

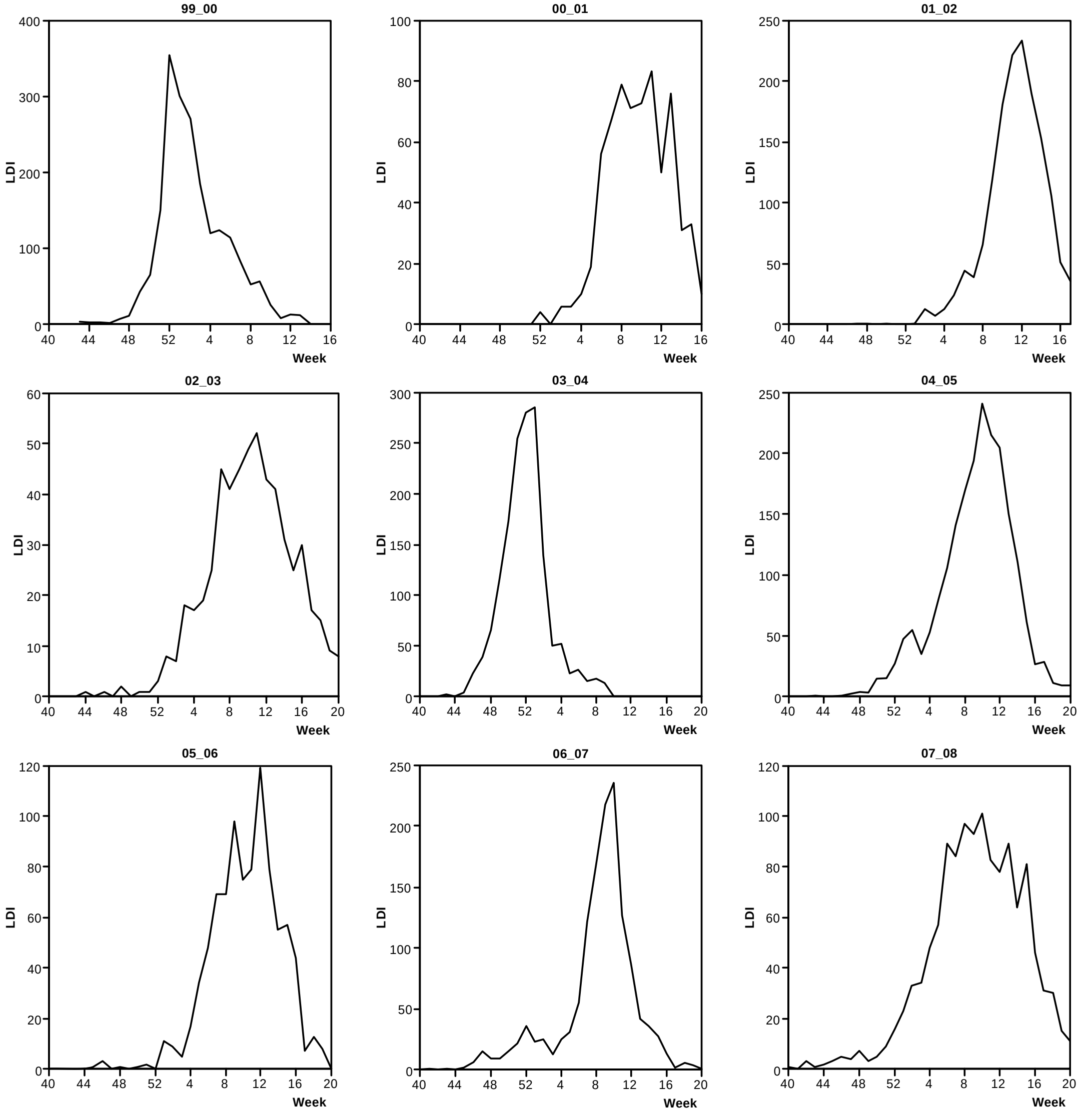

Figure 3. Number of laboratory diagnosed cases for the entire Sweden.

3.3.1. Non-epidemic period and outbreak

The number of weeks until the first laboratory diagnosed influenza case is shown in Table 6. There is a considerably variation between the years and also between laboratories. One reason for the latter may be differences in the size of the population. There may also be a difference in the incidence depending on other factors in the population, as well as differences in policies regarding testing. In general the largest cities, Stockholm, Göteborg, Malmö and Uppsala, are among the first with reported cases.

In Table 7 the time until the cumulative number of LDI is greater than 5 is shown. Here there is a clear tendency for the largest cities to be earliest. Since the population size of the different regions is unknown, a potential source of error is that the larger cities would reach a cumulative sum greater than five earlier than smaller cities.

Table 6. Number of weeks (since week 40) to the first laboratory diagnosed influenza case. The regions are sorted with respect to the median.

	99_00	**00_01**	**01_02**	**02_03**	**03_04**	**04_05**	**05_06**	**06_07**	**07_08**	**Median**
Göteborg	9	14	14	6	6	6	6	1	0	6
KS	3	14	7	8	5	7	11	10	4	7
HS	3	17	8	13	3	7	8	8	2	8
Umeå	3	17	15	12	7	10	5	3	8	8
Borås					6	13	17		6	9.5
Malmö	3	12	10	15	8	8	13	12	4	10
Lund									11	11
Uppsala	8	14	14	15	8	3	18	11	7	11
Skövde	8	14	15	4	5	13	16	8	14	13
Örebro	10	12	16	18	6	10	20	13	13	13
Falun	10	17	17	13	8	14	14	12	14	14
Halmstad	9	18	14	17	7	9	14	16	2	14
Helsingborg									14	14
Jönköping					11	24	14	19	9	14
Karlstad	6	19	14	15	8	11	17	12	17	14
Luleå	11			12	10	16	17		23	14
Sundsvall			14	20	8	16	16	11	13	14
Kristianstad							15	12	18	15
Kalmar	9	20	16	23	5	16	14	19	15	16
Karlskrona			16	16	7	19	13	15	17	16
Linköping	4	18	18	19	5	10	10	16	16	16
Uddevalla	11	16	16	19	7	9	17	16	11	16
Eskilstuna				15	7	18	24	22	15	16.5
Gävle			18	17	3	16	17	18	10	17
Växjö	12	18	16	18	7	25	17			17
Västerås	13	23	22	20	9	11	18	8	19	18
Östersund			19		20	14		18		18.5
Trollhättan							24	14		19
Median	9	17	15.5	15	7	12	16	12	13	

Table 7. Number of weeks (since week 40) until the cumulative number of LDI exceeds 5. The regions are sorted with respect to the median.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	Median
KS	7	17	17	13	7	11	16	16	8	13
HS	8	18	16	14	8	10	19	14	10	14
Göteborg	11	16	18	19	6	14	14	16	7	14
Malmö	9	17	18	18	9	13	18	14	14	14
Luleå	12				13	18	18			15.5
Falun	11	21	19	19	10	17	17	16	19	17
Lund									17	17
Skövde	9	18	17	17	6	17	16	11	18	17
Umeå	11	17	18	19	9	10	17	6	17	17
Uppsala	9	20	20	19	9	14	22	13	17	17
Eskilstuna					11	18			20	18
Helsingborg									18	18
Karlstad	9	24	19	21	9	16	18	15	22	18
Örebro	12	20	19	20	9	13	21	18	18	18
Sundsvall			17		10	19		19	18	18
Linköping	10		23	23	8	19	18	20	18	18.5
Gävle					11	21	25	19	19	19
Halmstad	11	20	21	22	8	19	19	18	14	19
Kalmar	12		19		11	19	23	24	22	19
Uddevalla	11	19	24	23	8	17	20	19	20	19
Västerås	14		24		10	17	23	21		19
Växjö	14	21	21	23	13					21
Borås					10	20	24		25	22
Jönköping					11	24	19	22	28	22
Karlskrona			27			22		22		22
Kristianstad							25	19	22	22
Trollhättan								22		22
Östersund			27			20				23.5
Median	11	19	19	19	9	17	19	18	18	

In Table 8 the correlation between the coordinates and the number of weeks until LDI exceed five is shown. None of the correlations were significant in it self. For longitude all but one correlation are negative. This is primarily caused by the early outbreak in Stockholm; by removing Stockholm all but two of the correlations changes sign. Our conclusion is that there is no strong relation between the coordinates and the time of outbreak.

Table 8. 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	Median
Latitude	-0.035	0.177	-0.126	-0.261	0.217	-0.129	-0.144	-0.348	0.007	-0.126
Longitude	-0.021	-0.046	-0.290	-0.431	0.291	-0.200	-0.003	-0.157	-0.133	-0.133

3.3.2. Peak

Both the time and the value of the peak vary from year to year, as can be seen in Tables 9 and 10. The variation of the time of the peak between years seems to be larger than the variation between laboratories, especially if we disregard the laboratories with only a few cases reported. The variation of the height of the peak is large both between years and laboratories. One factor in the variation among laboratories is that the size of the laboratories differs. There are also other factors such as population density, age distribution and the amount of travelling that could cause variation.

There seem to be no noticeable tendency for the larger cities to have an earlier peak.
No strong correlation between the time and height of the peak could be found. In general there seem to be a slight tendency that later peaks are lower.

Table 9. Number of weeks (since week 40) to the maximum value of LDI. The regions are sorted with respect to the median.

	99_00	**00_01**	**01_02**	**02_03**	**03_04**	**04_05**	**05_06**	**06_07**	**07_08**	**Median**
Lund									12	12
Luleå	13			12	13	17	18		23	15
Eskilstuna				15	13	18	24	22	18	18
Helsingborg									18	18
Gävle			18	17	11	22	27	20	19	19
Göteborg	13	18	21	20	11	19	19	22	19	19
Jönköping					11	24	19	22	9	19
Skövde	14	19	24	18	13	23	16	21	20	19
Borås					10	22	18		25	20
HS	12	25	24	19	12	22	20	21	20	20
Linköping	12	22	23	25	10	20	13	20	18	20
Östersund			27		20	20		18		20
Uddevalla	14	19	26	21	9	21	17	24	20	20
Falun	12	27	23	20	12	22	21	20	24	21
Halmstad	12	20	23	28	10	25	22	21	16	21
Karlstad	14	22	23	19	13	23	18	21	22	21
Umeå	13	23	24	22	13	17	22	12	21	21
Uppsala	14	21	22	23	13	21	24	22	20	21
Västerås	14	23	23	20	12	21	21	22	19	21
Växjö	19	21	23	21	13	25	17			21
Örebro		25	22	19	12	23	21	21	27	21.5
Karlskrona			27	16	7	22	27	22	17	22
Kristianstad							25	22	20	22
KS	12	19	23	22	12	22	24	21	22	22
Sundsvall			24	20	13	24	22	21	23	22
Trollhättan							24	21		22.5
Kalmar	14	27	25	23	11	22	23	19	25	23
Malmö	12	19	25	24	13	24	24	21	25	24
Median	13	21.5	23	20	12	22	21	21	20	

Table 10. The height of the peak of LDI as measured by the maximum value of LDI. The regions are sorted with respect to the median.

Laboratory	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	Median
Malmö	36	7	29	13	44	55	43	50	17	36
KS	70	20	35	11	54	33	18	25	28	28
Umeå	49	26	43	15	38	24	9	20	16	24
HS	59	21	28	10	37	22	20	27	21	22
Göteborg	21	13	8	5	16	9	17	32	18	16
Falun	15	6	26	7	42	22	7	17	8	15
Örebro	23	7	16	4	26	17	5	14	9	14
Skövde	23	13	33	6	29	40	7	13	4	13
Halmstad	22	5	9	3	8	11	8	12	4	8
Karlstad	30	4	8	3	8	11	4	10	3	8
Sundsvall			7	1	11	9	2	9	8	8
Uppsala	17	7	15	4	10	22	4	8	6	8
Kalmar	10	2	7	3	10	18	3	2	7	7
Jönköping					6	6	4	7	1	6
Uddevalla	18	5	13	3	11	8	4	5	3	5
Kristianstad							4	6	3	4
Linköping	10	2	8	4	7	7	2	4	3	4
Västerås	4	1	4	1	10	6	2	10	1	4
Borås					7	4	3		3	3.5
Eskilstuna				1	4	6	1	3	4	3.5
Luleå	6			1	4	3	4		3	3.5
Gävle			1	1	3	4	3	5	3	3
Helsingborg									3	3
Lund									3	3
Växjö	7	5	12	3	3	1	1			3
Karlskrona			2	1	1	4	2	3	1	2
Trollhättan							1	3		2
Östersund			3		1	3		1		2
Median	21	6.5	10.5	3	10	9	4	9	4	

3.4. Conclusions about the usefulness of LDI for outbreak detection

The data are complete (for the period 99_00 to 07_08) for more that half of the regions, including the largest cities (Table 5). 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.

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

4.1. Division into groups

In the tables above, we found that the large cities with good communications with other countries have a different pattern than the rest. This is also in accordance with the results in Bock and Pettersson (2006). We will use the same grouping, one group with the three metropolitan areas (Stockholm including Uppsala, Göteborg, and Malmö) and one with the rest of Sweden. Stockholm, Göteborg and Malmö all have populations considerably larger than the other cities, and are part of the metropolitan areas as defined by Statistiska centralbyrån (2005). Uppsala's population on the other hand is not much larger than the remaining cities, however the proximity and communications to Stockholm makes it suitable to include in the group. Also, the international airport Arlanda is situated about halfway between Stockholm and Uppsala.

Using Spearman's rank correlation we found that the laboratories in Stockholm, Göteborg and Malmö were highly correlated with each other, $\rho>0.7$ for most seasons. The correlation between Uppsala and the rest of the group were slightly lower, but still high enough to be reasonable to include in the group.

It could be argued that Lund and Borås also should be included in the metropolitan group due to their proximity to Malmö respectively Göteborg. The reporting from Borås is however inconsistent and Lund only has the 07_08 season. There are also other quality problems with the reports from these cities. We chose to exclude them from the metropolitan group.

We will denote the group with larger cities the metropolitan group and the other group the locality group. In Figure 4 the number of LDI cases for each group is shown. The total number of cases for the two groups is similar.

The similar number of cases for each of the groups facilitates the interpretation of the comparisons between the two groups. The total number of cases for all seasons, up to the peak was 3379 for the metropolitan group and 3205 for the locality group.

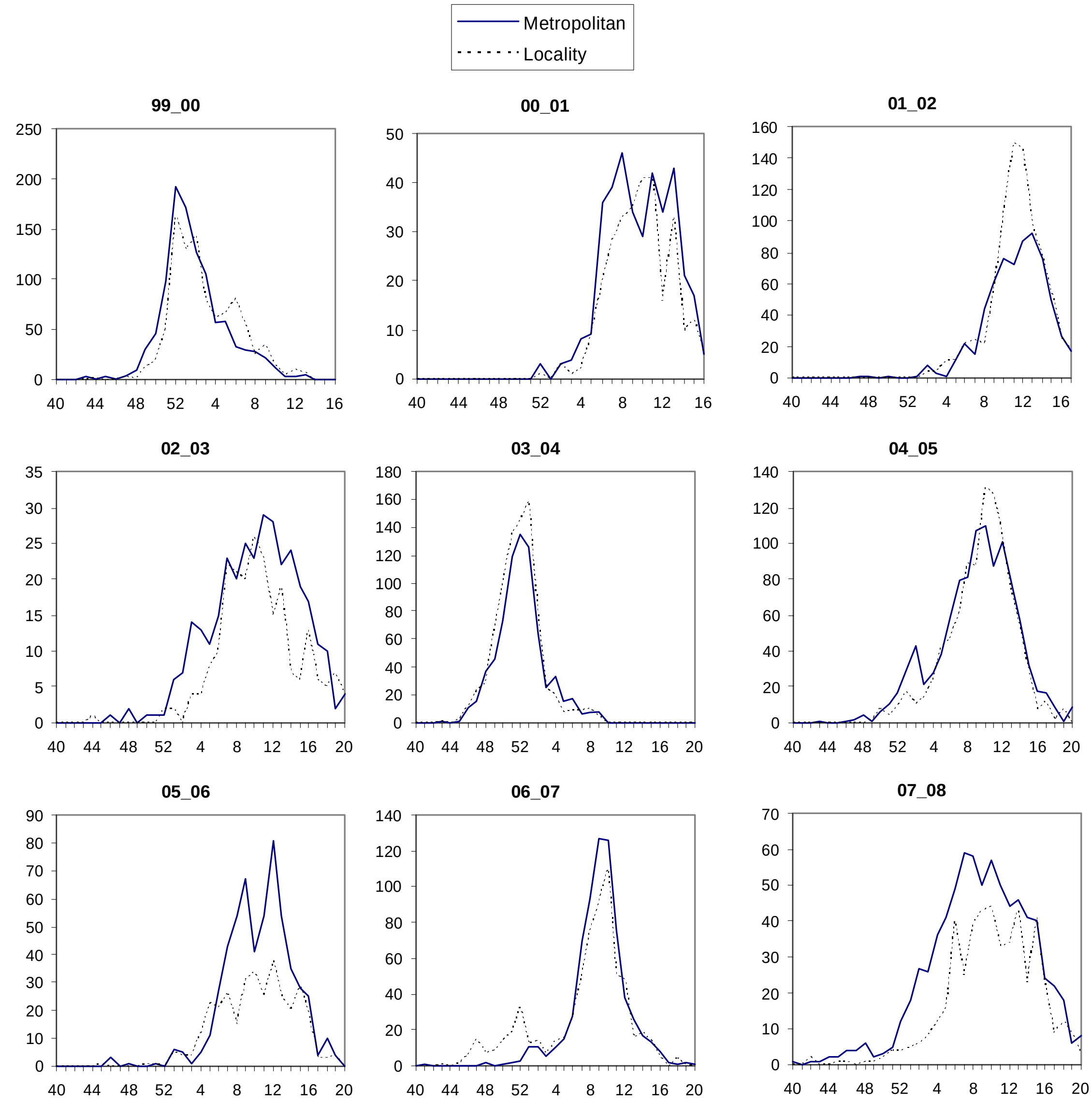

Figure 4. Number of laboratory diagnosed cases for the metropolitan group, Stockholm/ Uppsala, Göteborg, and Malmö (solid line) and the locality group, the rest of Sweden (dotted line).

4.2. Differences in time of start of increase of incidence

In Table 11 the number of weeks until the cumulative number of LDI cases exceeds 9 is shown. The metropolitan group is earlier for most years. In the 06_07 season an early outbreak in Umeå caused the locality group to be earlier, while the remaining cities in the group did not have a single case. The median of the differences is 2.

Table 11. Number of weeks until the cumulative number of LDI exceeds 9.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08
Locality	9	17	16	16	6	10	14	7	11
Metropolitan	7	15	14	13	6	10	13	13	6
Difference	2	2	2	3	0	0	1	-6	5

Table 11 suggests that there is a time lag between the two groups. Additional analysis with Spearman's rank correlation was performed on each season to test if a lag would increase the correlation between the groups. Using data from the entire season, the correlation without a lag was high, around 0.9 for all season except 06_07. A lag of two weeks only increased the correlation for one season.

The same analysis of the start of the season up to the peak showed a small increase in correlation for four seasons with a lag of two weeks. For the remaining seasons the correlation decreased slightly.

Concentrating on the outbreak, taking only the observations from the start up to until the total number of observed cases had exceeded 30 gave an increase in correlation for the same four seasons as up to the peak. The difference between using a lag or not was however larger. The results are shown in Table 12. A lag of two weeks gives the largest median over the years. The median is nearly as high for a lag of three weeks but there the range is larger.

Table 12. Spearman's rank correlation between the metropolitan and locality groups for different seasons. All observations from the start until the total number of LDI exceeded 30 was used for lag zero. Later weeks was added to the locality group to get corresponding lagged values.

		Season										
		99_00	**00_01**	**01_02**	**02_03**	**03_04**	**04_05**	**05_06**	**06_07**	**07_08**	**Median**	**Range**
Lag	**0**	0.62	0.98	0.79	0.75	0.99	0.80	0.74	0.55	0.72	0.75	0.44
	1	0.69	0.76	0.83	0.78	0.72	0.84	0.51	0.42	0.66	0.72	0.42
	2	0.84	0.93	0.75	0.84	0.86	0.92	0.71	0.35	0.77	0.84	0.58
	3	0.69	0.86	0.70	0.81	0.86	0.92	0.67	0.09	0.86	0.81	0.83

4.3. Difference in evidence of increase

In Frisén and Andersson (2007) a semi-parametric method of surveillance is applied to the Swedish LDI data for the entire Sweden. Figure 5 shows the alarm statistic of the method applied to the two groups. The metropolitan group has a tendency to rise earlier than the locality group, and thus can be expected to give an alarm or an early warning earlier.

4.4. Slope of the expected incidence

Due to the interaction between estimates of the start of the outbreak and the slope of outbreak it is difficult to determine the slope by ordinary estimation of parametric curves. Andersson et al. (2008) use the time difference between the (interpolated) time when the total LDI in Sweden exceeds 30 and 10 as an indicator of the slope. Since each of the groups is about half the size of the total we used the difference between 15 and 5. It was suggested in Andersson et al. (2008) that smoothing by unimodal regression could be used to reduce some of the random variation in the available data without assumption of a parametric model. Using these techniques we found no significant difference between the slopes of the two groups.

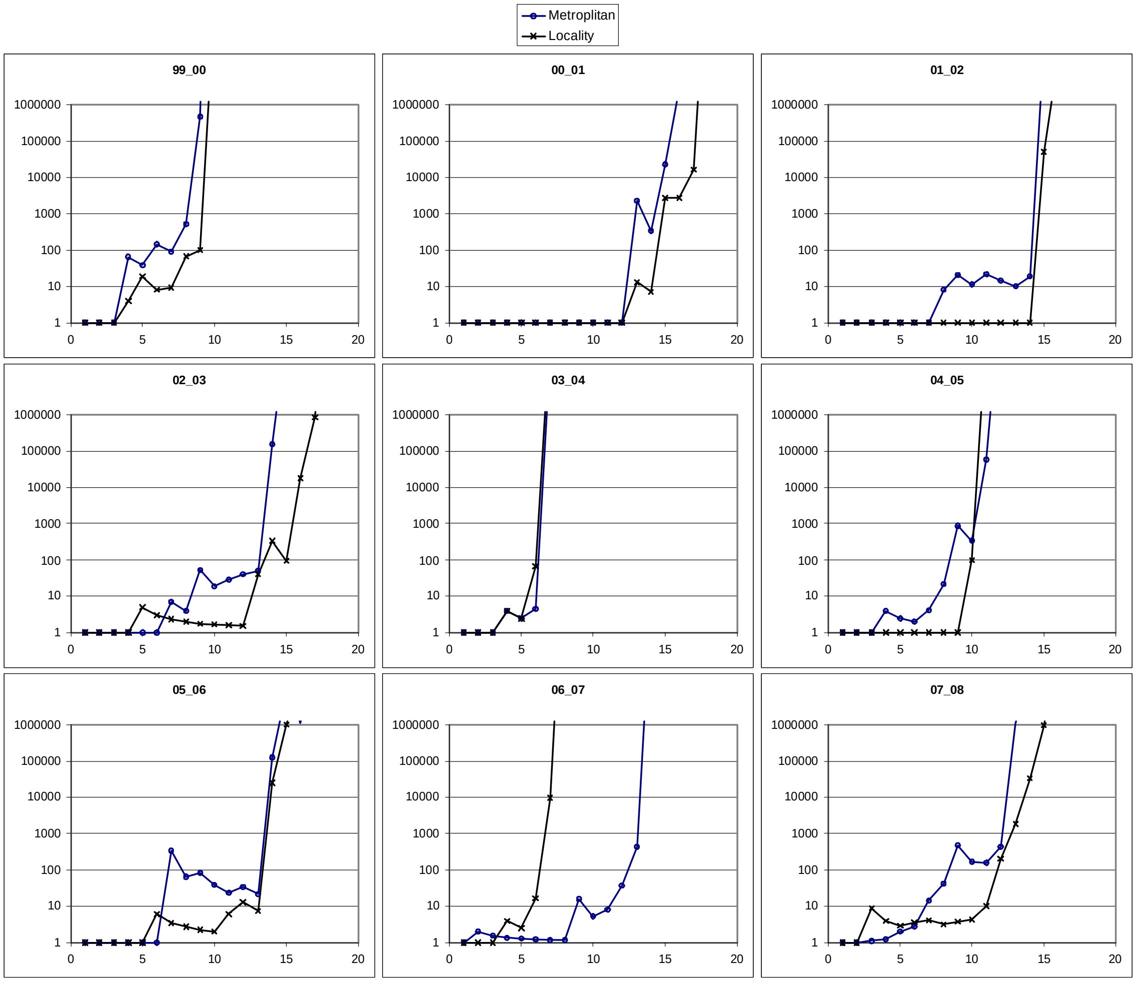

Figure 5. OutP alarm statistics for the groups. The dots represents the metropolitan group and the crosses represent the locality group.

4.5. Parametric models of the expected incidence

In order to make a simulation study of the properties of a method of surveillance some sort of parametric model is needed. In Frisén and Andersson (2007) the model

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

is used for a typical curve for the total LDI 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 for the season 2003-2004 which was neither a very severe or very mild outbreak. The estimates of the parameters were $\beta_0 = -0.26$ and $\beta_1 = 0.826$.

The locality and metropolitan groups each have about half the number of cases as the total. The above curve thus has to be divided by two to represent the same pattern as the total. This curve fitted well to the data for the same season (2003-2004) for some values of the starting time. It also fitted rather well for some other seasons while a good fit to all seasons could not be expected due to the marked differences between the seasons.

5. Concluding remarks

The spatial ILI data for the last years had several major deficiencies. Thus, we had to base our conclusions for ILI on data for earlier years, which also had quality problems. Underreporting is a major problem. It is most evident in the beginning of the season, which we were primarily interested in. One possible explanation is that the number of reported cases may be lower due to physicians' expectation that the influenza season has not started yet. There is also a considerable decrease in the number of visiting patients during the Christmas holiday.

LDI data also had quality problems, but these were not as severe as those of ILI. Thus, we based our conclusions about spatial patterns on LDI data.

The number of reported cases is relatively low for both ILI and LDI. During some years, some regions had only a few cases. Hence, there is a need for aggregation of data. It is important that the spatial differences are not removed by the aggregation. We examined some natural spatial patterns such as those based on geographical coordinates. 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 and Malmö, the onset of the influenza outbreak seems to occur earlier than in the rest of the country. These regions all have major international airports nearby (Arlanda, Landvetter and Kastrup), and commuting is common. Furthermore, the population density is higher here than in the rest of the country.

Comparisons between the metropolitan and locality groups by the time a certain incidence was reached, by the correlation between lagged variables, and by graphs of the incidence of the onset of the outbreak indicated that for the metropolitan group, the onset of the outbreak came about two weeks earlier than for the locality group. As for the incidence slope at the onset, no evidence was found for a difference between the two groups.

Acknowledgements

The author is grateful to Professor Marianne Frisén and Associate professor Eva Andersson for supervision of this work. Ph. Licentiate Kjell Pettersson has given many constructive comments. The authors of Bock and Pettersson (2006) have made their data and analyzes available. Sandra Rubinova at the Swedish Institute for Infectious Disease Control has given expert information about the data used. The research was supported by the Swedish Emergency Management Agency (grant 0622/204).

References

Andersson, E., Bock, D., and Frisén, M. (2007). Modeling influenza incidence for the purpose of on-line monitoring. *Statistical Methods in Medical Research*.

Andersson, E., Kuhlmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S., and Frisén, M. (2008). Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden. *Scandinavian Journal of Public Health* **36**, 475-482.

Bock, D., Andersson, E., and Frisén, M. (2008). Statistical surveillance of epidemics: Peak detection of influenza in Sweden. *Biometrical Journal* **50**, 71-85.

Bock, D., and Pettersson, K. (2006). Exploratory analysis of spatial aspects on the Swedish influenza data. In *Smittskyddsinstitutets rapportserie*. Stockholm: 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* **3**, e401.

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

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

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

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.

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* **6**, 58.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society A* **164**, 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* **3**, e151.

Sonesson, C. (2007). A CUSUM framework for detection of space-time disease clusters using scan statistics. *Statistics in Medicine* **26**, 4770-4789.

Statistiska centralbyrån (2005). Geografin i statistiken - regionala indelningar i Sverige In *MIS*, 42: Statistiska centralbyrån.

Szucs, T. (1999). The socio-economic burden of influenza. *Journal of Antimicrobial Chemotherapy* **4**, 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* **35**, 3-12.

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

Research Report

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.
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.