# Influenza Forecast: Comparison of Case-Based Reasoning and Statistical Methods

Tina Waligora and Rainer Schmidt

Universität Rostock, Institut für Medizinische Informatik und Biometrie,
Rembrandtstr. 16 / 17, D-18055 Rostock, Germany
{tina.waligora,rainer.schmidt}@medizin.uni-rostock.de

**Abstract.** Influenza is the last of the classic plagues of the past, which still has to be brought under control. It causes a lot of costs: prolonged stays in hospitals and especially many days of unfitness for work. Therefore many of the most developed countries have started to create influenza surveillance systems. Mostly statistical methods are applied to predict influenza epidemics. However, the results are rather moderate, because influenza waves occur in irregular cycles. We have developed a method that combines Case-Based Reasoning with temporal abstraction. Here we compare experimental results of our method and statistical methods.

## 1 Introduction

Since influenza results in many costs, e.g. for delayed stays in hospital and especially for an increased number of unfitness for work, many of the most developed countries have started to generate influenza surveillance systems (e.g. US: www.flustar.com, France [1], and Japan [2]). The idea is to predict influenza waves or even epidemics as early as possible and to indicate appropriate actions like starting vaccination campaigns or advising high-risk groups to stay at home.

Mostly statistical methods are applied to predict influenza epidemics. However, the results are rather moderate, because influenza waves occur in irregular cycles and Farrington pointed out that statistical methods have difficulties to cope with infectious diseases characterised by irregular cyclic behaviour [3].

So, we have developed a method that combines Case-Based Reasoning with temporal abstraction. Before we explain our method and subsequently present comparative results of our method and of statistical methods, we discuss the question which data should be used.

### 1.1 Data

It is well known that a couple of factors can be responsible for influenza outbreaks. One of them is the weather. It is often assumed that a strong winter increases the spread of influenza. However, studies could reveal only an extremely small relation between temperatures and spread of influenza [4]. Dowell even suggests that the occurrence of influenza in winter is not related to the temperature but to the annual light/dark pattern [5]. Other influences are the mutations of the virus and influenza outbreaks in foreign countries, even as far as Hongkong. Unfortunately, no exact knowledge about these influences is available. So far, only seasonal behaviour of influenza is well known from observations [3, 5]. Therefore all surveillance systems focus on observed numbers of infected people, especially on their increase.

J.L. Oliveira et al. (Eds.): ISBMDA 2005, LNBI 3745, pp. 202–210, 2005.

Considerations about how to forecast influenza begin with questions about which data should be used and which data are available. The answer varies with the country and its health organisation.

In some countries with rather private health systems like Germany, research groups interested in predicting influenza have started to develop surveillance nets based on voluntary participation of general practitioners. They felt that the data collected and provided by official health centres were insufficient, because they are usually available with a delay of two or even three weeks. These surveillance nets are based on general practitioners who give once a week some sort of standardised reports. It needs a huge effort to initiate and organise such nets and rural areas are very often not adequately represented, because it is difficult to find doctors willing to participate. Since the reports are always subjective, misjudgements and data interpretation errors may occur, which may lead to false assessments – especially in areas with low density of participating doctors.

The alternative means to use official data from health centres. In Germany, these data are more objective, because they contain reports and laboratory results of all occurrences of notifiable diseases. Unfortunately, because of the hierarchical and bureaucratic organisation of the health centres, the availability of these data is delayed for at least two weeks. In countries with more public health systems, sometimes the situation seems to be much better, e.g. in Japan [2].

However, we have chosen another alternative. Since 1997 we receive data for our federal state Mecklenburg-Western Pomerania from the main health insurance scheme. These data are sick certificates of employees and of people who receive unemployment benefit. Fortunately we get the data daily. Of course there is a short delay between doctors writing the certificates and the insurance scheme receiving them by mail from their policyholders. We do not recur on the days when the certificates have been issued by doctors, but on the daily data sets received by the insurance scheme. Since there are some daily fluctuations by chance, influenza surveillance systems usually use weekly aggregated data.

The disadvantage of using insurance data is their superficiality, because the certificates usually contain just the first diagnoses, which might be refined or changed later on. However, for influenza this is only a minor problem, because the symptoms of influenza, acute bronchitis, etc. are so similar that most surveillance groups use a superficial category anyway, namely all acute respiratory diseases to infer influenza.

## 2 Prognostic Methods

All influenza surveillance systems make use of developments in the past. Most of them have tried statistical methods. The usual idea is to compute mean values and standard courses based on weekly incidences of former influenza seasons (from October till March) and to analyse deviations from a statistic normal situation.

Influenza waves usually occur only once a season, but they start at different time points and have extremely different intensities. Since Farrington pointed out that statistical methods are inappropriate for diseases like influenza that are characterised by irregular cyclic temporal spreads [3], we have developed a method that uses former influenza seasons more explicitly. We apply the Case-Based Reasoning idea: that means to determine the most similar former courses of weekly incidences and to use them to decide whether a warning is appropriate. Viboud [6] from the group that is responsible for the French Surveillance net has developed a method that is very similar to our one. However, both methods differ in their intentions. Viboud attempts to predict incidences few weeks in advance, while we are interested in more practical results, namely in the computation of appropriate warnings.

### 2.1 Case-Base Reasoning to Forecast Influenza

Inspired by our former program for the prognosis of kidney function courses [7], we have developed a method to decide about the appropriateness of warnings against approaching influenza waves (figure 1).

Every influenza season consists of 26 weeks (from October till March). Since we consider weekly incidences, seasons are represented as sequences of 26 numeric values. Each week it has to be decided anew whether a warning is appropriate or not. For this decision, just the recent development is important. So, we consider only a sequence of the four most recent weeks. When an influenza season is finished, it is separated into 23 four-week courses; all of them are stored as cases in the case base.

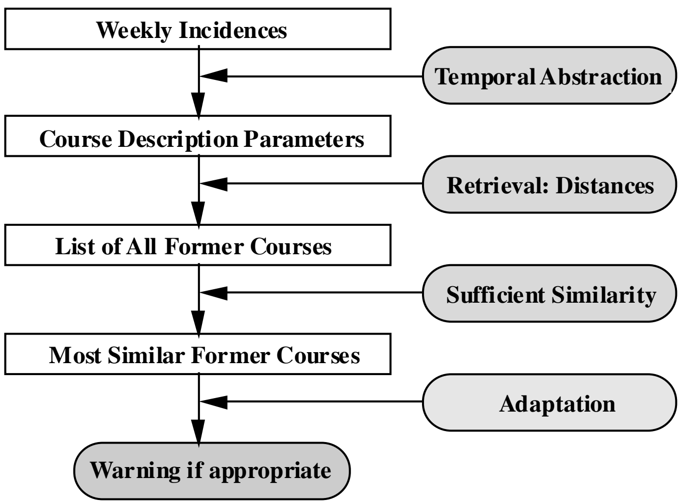


**Fig. 1.** Case-Based Reasoning method to forecast influenza

The first step of our method is a temporal abstraction of a sequence of four weekly incidences into three trend descriptions that assess the changes from last week to this week, from last but one week to this week and so forth. Secondly, these three assessments and the four weekly incidences are used to determine similarities between a current query course and all four-weeks courses stored in the case base. Our intention for using these two sorts of parameters is to ensure that a query course and an appropriate similar course are on the same level (similar weekly incidences) and that they have similar changes on time (similar assessments). More details about these first two steps of our method can be found in [8].

The result of computing distances is a very long list of all former four-week courses sorted according to their distances in respect to the query course. For the decision whether a warning is appropriate, this list is not really helpful, because most of the former courses are rather dissimilar to the query course. So, the next step means to find the most similar ones. We decided to filter the most similar cases by applying two explicit similarity conditions. First, the difference concerning the sum of the three trend assessments between a query course and a similar course has to be below a threshold X. This condition guarantees similar changes on time. And secondly, the difference concerning the incidences of the current weeks must be below a threshold Y. This second condition guarantees an equal level of the current week of a similar case and the current week of the query course. We have learned good settings for the threshold parameters X and Y by taking in turn one season out of the case base and comparing the results when varying the settings.

The result of this third step usually is a very small list containing only the most similar former courses. As in compositional adaptation [9] we take the solutions of a couple of similar cases into account, namely of all courses in this small list.

In retrospect, we have marked those time points of the former influenza seasons where we believed a warning would have been appropriate; e.g. in the 4th week of 2001, which is the 17th week of the 2000/2001 season (marked as square in fig.2).

For the decision to warn, we split the list of the most similar courses in two lists. One list contains those courses where a warning was appropriate; the second list gets the other ones. For both of these new lists we compute their sums of the reciprocal distances of their courses to get sums of similarities. Subsequently, the decision about the appropriateness of a warning depends on the question which of these two sums is bigger.

## 2.2 Statistical Methods to Forecast Influenza

A couple of statistical tests are available. Under the assumption of binomial distribution, we have tested whether the observed weekly count of infected people is significant. With the following formula the probability of exactly k insured people being infected:

$$P(X=k)=\binom{n}{k} p^k (1-p)^{n-k} \, for \, k=0,\dots,n$$

where

k = number of observed insured people being infected

n = number of insured people

and

$$p=\frac{Number\ of\ infected\ people\ on\ average}{Number\ of\ insured\ people}$$

Here, “on average” means the average number of infected people per week concerning the specific months. Concerning the whole time period of three years or just the influenza seasons would be too vague, because influenza mainly occurs in some months, and concerning specific calendar week might be randomly.

If the sum of the probabilities of k people being infected, k+1 people being infected etc. is 5% or more, an influenza wave can be assumed:

$$P(X \geq i)=\sum_{k=1}^{n} P(X=k) \, for \, i=0,\dots,n$$

# 3 Experimental Results

We have performed some experiments to compare both methods. However, we used different data. For doctors it is extremely difficult to distinguish between a real influenza infection and other acute respiratory infections, especially at the first diagnosis. So, in most influenza surveillance systems influenza is usually inferred from the counts of all acute respiratory infections.

For the Case-Based Reasoning method we used acute respiratory infections (ICD9: 460 to 487 and ICD10: J00 to J99, except very few chronic diseases) for the influenza seasons from 1997 to 2002.

For the statistical tests we just used data of real influenza infections (ICD10: J10 and J11) for three years, namely for 2000, 2001, and 2002, because the ICD10 code was introduced in 2000 and for 1997-1999 (ICD9: 487) the counts did not fit together the later ICD10 counts.

For both experiments, we have not used more up-to-date data, because in Mecklenburg-Western Pomerania the last influenza wave occurred in early 2001 and the course of the 2001/2002 season is typical for the following ones.

## 3.1 Case-Based Reasoning Method

First, we have marked those time points where we, in retrospect, believed a warning would have been appropriate (the three squares in figure 2). Later on we assumed that these warnings might be a bit late. So, we have additionally attempted earlier desired warnings (the three circles in figure 2).

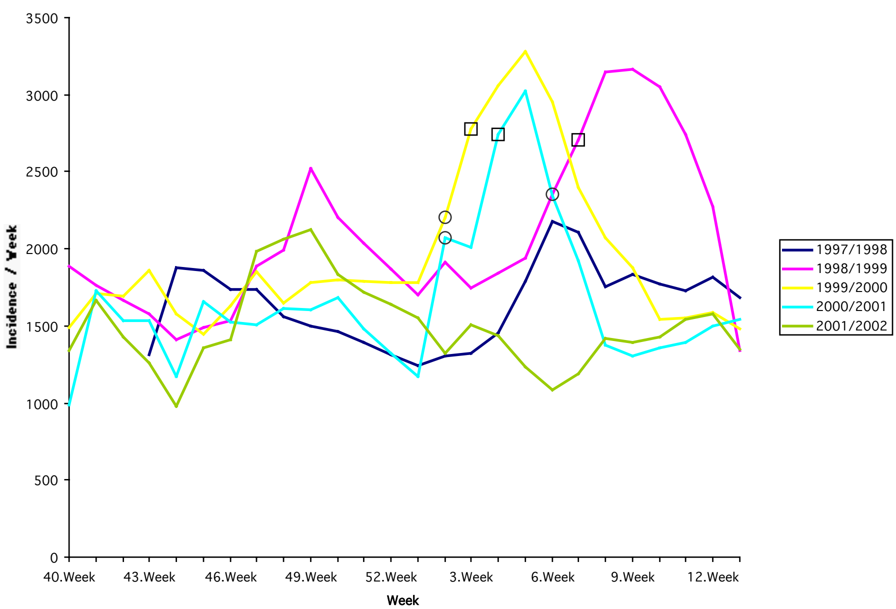


**Fig. 2.** Influenza seasons of Mecklenburg-Western Pomerania, ICD10: J00-J99, ICD9: 460-487

### 3.1.1 First Experiment

For our first tests, we used the five seasons shown in figure 2 with the desired warnings depicted as squares. In turn we used one season as query course. Furthermore, we wanted to discover how much the results are improved by the number of seasons stored in the case base. So, for every query season we varied the case base, and we did not only put the seasons in chronological order into the case base, but attempted every combination. That means, for each query season we made four attempts with one season in the case base, six attempts with two seasons etc.

The results are shown in table 1. Sensitivity means proportion of computed warnings to desired warnings; specificity means proportion of computed "non-warnings" to desired "non-warnings".

**Table 1.** Sensitivity and specificity of our first experiment

| | Sensitivity | Specificity |
|---|---|---|
| 1 season in case base | 50 % | 100 % |
| 2 seasons in case base | 83 % | 100 % |
| 3 seasons in case base | 100 % | 100 % |
| 4 seasons in case base | 100 % | 100 % |

At first glance the results seem to be very good: there are no false warnings and to exactly compute the desired warnings, for every season it is sufficient to use just three of the four remaining seasons as case base. However, since for every query season 23 decisions have to be made, most of them are obvious "non-warnings", a few are follow-up warnings (determined by a simple heuristic when the week before a warning or a follow-up warning was computed), and only few decisions are really crucial.

### 3.1.2 Earlier Warnings

Since we imagined that the desired warnings of our first experiment might be a bit late, we tried earlier ones in a second experiment, in figure 2 depicted as circles. We made the same experiment again and the results are shown in table 2.

**Table 2.** Sensitivity and specificity of our second experiment: with earlier warnings

| | Sensitivity | Specificity |
|---|---|---|
| 1 season in case base | 45 % | 95 % |
| 2 seasons in case base | 69 % | 96,7 % |
| 3 seasons in case base | 80 % | 97,2 % |
| 4 seasons in case base | 80 % | 96,1 % |

Now it is more difficult to compute the new desired warnings. However, the problems are mainly caused by the peak in the $49^{th}$ week of 1998, which is the $10^{th}$ week of the 1998/1999 season. Since the incidences of this peak are higher than the incidences of the desired warnings and the developments are similar too, consequently a warning is computed. Only in retrospect it becomes clear that this was not the beginning of an influenza wave. And since this peak is marked as not worth for a warning, it prevents our program from computing desired warnings for other seasons.

However, this is not so much a problem of the method, but rather a question of the availability of appropriate data. Since we use health insurance data, we do not have access to laboratory results, which often indicate causes. In fact, concerning the analysis of such data by the Robert-Koch Institute [11], the peak in the 49th weak of 1998 was probably (but this was never definitely proved) the result of a pathogen (respiratory syntactical virus) that causes similar symptoms as influenza. Unfortunately, such data from health centres even the Robert-Koch Institute gets only delayed (about two weeks).

## 3.2 Statistical Tests

For the statistical tests we did not consider all acute respiratory infected people but only those with real influenza infections (ICD10: J10 and J11). Since the counts of ICD9 and ICD10 did not fit together, we used only data for 2000, 2001, and 2002 (figure 3).

Though the considered infections differ, in both figures (see figure 2 and figure 3) the increase of seasons 2000 and 2001 occurs obviously at the same moment, namely in the second calendar week. The fact that influenza waves start in two following seasons in the same week is poor chance. Sometimes influenza waves start much later, see e.g. the 1998/1999 season (figure 2). We performed tests under the assumption of binomial distribution (see section 2.2). Parts of results, namely for the first four calendar weeks, are shown in table 3. A significant count of infected people is indicated, when the value is below 5%.

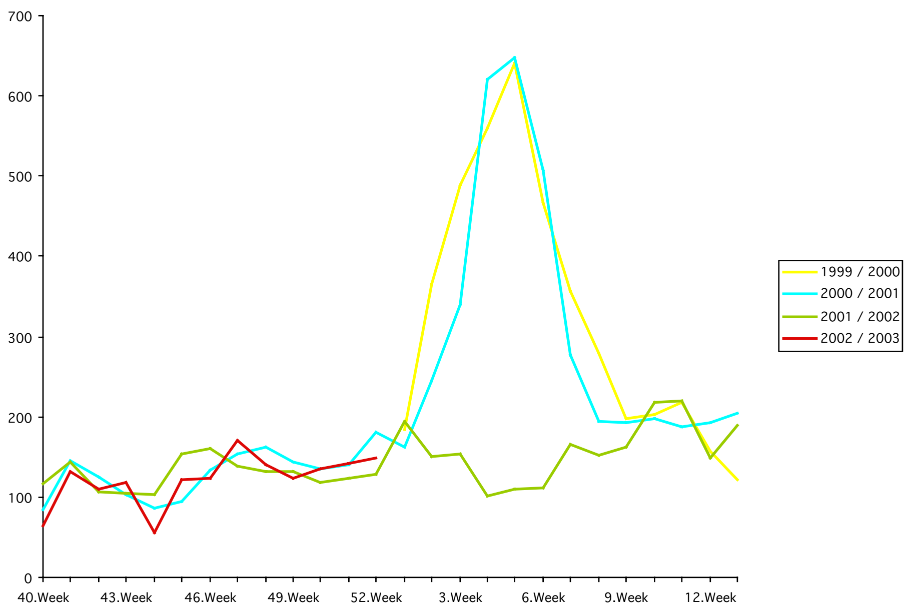


**Fig. 3.** Influenza seasons of Mecklenburg-Western Pomerania, ICD10: J10+J11

For 2000 table 3 shows the desired result, namely the beginning of an influenza wave in the second calendar week. For 2001 the obviously desired result is a start of an influenza wave in the second week too, but the statistical test discovers the start not until the fourth week. Of course, one problem is that we considered just three seasons. However, the results illuminate a problem of statistical methods applied to diseases with cyclical behaviour. In two of the three considered seasons influenza waves occurred which increase the average value very much. So, instead of using a general average value only those weeks without influenza waves should be considered as "normal" and should be used for computing average values. However, sometimes it is difficult to decide whether a week is "normal" and for specific calendar weeks there may be just very few "normal" ones.

**Table 3.** Results of tests under assumption of binomial distribution. NIL means that no computation was necessary, because the observed count is even below the average value

| | Counts 2000 | % | Counts 2001 | % | Counts 2002 | % |
|---|---|---|---|---|---|---|
| 1.Week | 184 | 41 | 163 | Nil | 195 | |
| 2.Week | 366 | 0,1 | 246 | Nil (32) | 151 | Nil |
| 3.Week | 489 | 0,1 | 340 | 23 | 154 | Nil |
| 4.Week | 559 | 0,1 | 620 | 0,1 | 102 | Nil |

# 4 Conclusion

In contrast to most medical diagnostic problems, we cannot ask experts about the correctness of the computed warnings. Instead, nobody knows in which week a first warning should be computed. However, the simultaneous increase in both sorts of data indicates that for the Case-Based Reasoning method the earlier warnings of the second experiment are probably ideal moments for first warnings against approaching influenza waves.

So far, it is difficult to assess the quality of our Case-Based Reasoning method. However, the results are at least as good as with statistical tests that assume binomial distribution. Unfortunately for our research, there has not occurred an influenza wave since 2002 in Mecklenburg-Western Pomerania.

Furthermore, we believe that the results of influenza surveillance depend more on the data than on the method. This does not only mean the a priori quality of the data and the speed of their availability, but additionally the quality for discriminating risky situations. The a priori quality of our health insurance data is rather poor, especially the diagnoses are often superficial, but there is only a very short delay concerning their availability. Official data from German health centres are more profound, but for bureaucratic reasons there availability is delayed for too long.

# References


1. Prou, M., Long, A., Wilson, M., Jacquez, G., Wackernagel, H., Carrat, F.: Exploratory Temporal-Spatial Analysis of Influenza Epidemics in France. In: Flahault, A., Viboud, C., Toubiana, L., Valleron, A.-J. (eds.): Abstracts of the 3rd International Workshop on Geography and Medicine, Paris, October 17-19 (2001) 17
2. Shindo, N. et al.: Distribution of the Influenza Warning Map by Internet. In: Flahault, A., Viboud, C., Toubiana, L., Valleron, A.-J. (eds.): Abstracts of the 3rd International Workshop on Geography and Medicine, Paris, October 17-19 (2001) 16

3. Farrington, C.P., Beale, A.D.:,The Detection of Outbreaks of Infectious Diseases. In: Gierl L. et al. (eds.): International Workshop on Geomedical Systems, Teubner, Stuttgart (1997) 97-117
4. Rusticucci, M., Bettolli, M.L., De Los Angeles, Harris, M., Martinez, L., Podesta, O.: Influenza outbreaks and weather conditions in Argentina. In: Toubiana, L., Viboud, C., Flahault, A., Valleron, A.-J. (eds.): Geography and Health, Inserm, Paris (2003) 125-140.
5. Dowell, S.F.: Seasonal Variation in host susceptibility and cycles of certain infectious diseases. Emerg Inf Dis 7 (3) (2001): 369-374
6. Viboud, C. et al.: Forecasting the spatio-temporal spread of influenza epidemics by the method of analogues. In: Abstracts of 22nd Annual Conference of the International Society of Clinical Biostatistics, Stockholm, August 20-24 (2001) 71
7. Schmidt, R., Gierl, L.: Prognoses for Multiparametric Time Courses. In: Brause, R.W., Hanisch, E. (eds.): Medical Data Analysis. Proceedings of ISMDA 2000, Springer-Verlag, Berlin (2000) 23-33
8. Schmidt, R., Gierl, L.: A prognostic model for temporal courses that combines temporal abstraction and case-based reasoning. Int J Medical Informatics 74 (2005) 307-315
9. Wilke, W., Smyth, B., Cunningham, P.: Using Configuration Techniques for Adaptation, In: Lenz, M., et al. (eds.): Case-Based Reasoning Technology, From Foundations to Applications. Springer-Verlag, Berlin (1998) 139-168