Statistics in Medicine

A Statistical Derivation of Thresholds for Bioterrorism Monitoring of Daily Cases of Flu-Like Symptoms

Journal:	*Statistics in Medicine*
Manuscript ID:	draft
Wiley - Manuscript type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Miller, Austin; University at Buffalo, Biostatistics Carter, Randolph; University at Buffalo, Biostatistics
Keywords:	Disease surveillance, Threshold model, Poisson regression, Prediction limits

A Statistical Derivation of Thresholds for Bioterrorism Monitoring of Daily Cases of Flu-Like Symptoms

Austin Miller*† and Randy L. Carter

Department of Biostatistics, State University of New York at Buffalo, Buffalo, NY, U.S.A

SUMMARY

The UB Population Health Observatory developed a Bioterrorism trigger that can alert the Erie County Department of Health when the observed count of flu-like-symptoms cases on a particular day exceeds a threshold estimated by a statistical model. Daily count of flu-like symptoms cases between Aug 1, 1999 and Nov 30, 2002 were divided into training and model validation subsets. Explanatory variables were restricted to seven lags of daily temperature and precipitation each. These readily available data ensure easy implementation and updating of the model. Indicator variables for week-day and month were also included in the model. Upper $100(1-\alpha)\%$ prediction limits were calculated for each observation using Monte Carlo simulation. The prediction limits served as daily thresholds. Assuming no unusual events impacted the number of flu cases during the training and validation time frame, the 95% prediction limits can be expected to generate about 32 false alarms per year. The $100(1-\alpha)\%$ prediction limit can be adjusted to control the number false alarms. The uniqueness of this methodology lies in distributional assumptions of the residuals, necessitating the simulation.

KEY WORDS: Disease surveillance; Threshold model; Poisson regression; Prediction limits;

1. INTRODUCTION

Many potential bioterrorism agents and devastating emerging diseases, such as SARS or the avian flu, manifest themselves with flu-like symptoms. This fact has prompted interest in the development of real-time monitoring systems for early detection of excess numbers of influenza like illnesses (ILI). For example, the Erie County Department of Health and the Calspan University at Buffalo Research Center (CUBRC) collaborated to plan a syndromic surveillance system for Erie County, New York. The proposed surveillance system was extensive, encompassing input from all emergency services, including hospital emergency rooms, 911 calls and other first responders. The current

* Correspondence to: Austin Miller, Department of Biostatistics, State University of New York at Buffalo, Farber Hall Room 264, 3435 Main Street, Buffalo, NY 14214-3000, USA
† E-mail: am65@buffalo.edu

research was motivated by the need to incorporate statistically based early warning alerts of excess numbers of ILI cases into such surveillance systems.

Three general approaches to statistical surveillance systems have been used: (1) those that are based on Markov models; (2) change point detection methods; and (3) those that are based prediction thresholds.

Methods based on hidden Markov models were proposed by Le Strat [1] and Rath [2], for example. Generally, these approaches use a training data set to derive an algorithm for classifying future observed counts into normal (baseline) or epidemic states. When the goal is surveillance for bioterrorism or emerging diseases, the epidemic state usually has not occurred in the training data set. Thus, Markov model approaches are not applicable.

Hutwagner [3] described the CDC Early Aberration Reporting System (EARS) which employs CUSUM change-point detection and other methodologies. Rossi [4] extended the CUSUM method to accommodated covariates. Rolfhamre [5] found CUSUM methods to be less sensitive than other methods, but Cowling [6] found parity in the effectiveness of CUSUM and time series approaches. Farrington [7] argued that maintenance requirements make CUSUM methods inappropriate.

Prediction threshold methods for identifying significantly larger than expected numbers of cases of an event of interest are represented extensively in the literature. Three general types of threshold methods have been used: that is, those based on 1) parametric regression models without regard for historical counts [7,8,9]; 2.) smoothing methods [10]; and 3.) time series models (*e.g.*, ARIMA models could be used; such as those studied by Zeger [11] and Watier [12]; or branching processes as discussed by Diggle, *et al.* [13] and extended by Held [14]).

We shall employ a first order autoregressive model with daily weather variables and indicator variables for day of the week and month of the year to define thresholds for early identification of excess numbers of ILI cases. An association between weather and flu incidence has been recognized by many. The causal link, however, is disputed. In 1920, Huntington [15] tracked weekly deaths from influenza in New York City from 1889 through 1918, noting the “marked effect” of low temperature on increasing the incidence of influenza. Schulman and Kilbourne [16], however, observed that mice were more likely to get sick in the winter in spite of strictly controlled temperature and humidity. Several reasons for the observed association have been postulated. Season may impact immune system resistance to the virus. Human reactions to the live virus vaccines are more frequent in winter than in summer [17]. In cold weather, indoor crowding becomes more common, and buildings are less well ventilated. Spread of influenza is greatly influenced by the density and mass of a population [18]. Influenza spreads exceptionally well in concentrated communities [19]. Regardless of the cause, the association between weather and number of ILI cases provides information that, if utilized, should improve the accuracy with which we identify excess numbers of ILI cases.

Previous authors typically have used normal theory quantiles in their derivations of thresholds. Models used for prediction, however, contain a model error whose distribution is a centered version of the conditional distribution of the outcome variable (Y=count), which typically is not normally distributed given the covariates. The normality assumption, therefore, is heuristic. It is more natural to assume that counts have a Poisson distribution. It is common, however, to observe extra Poisson variation in observed counts when employing Poisson models, perhaps because of the association of ILI incidence with weather, or other variables that vary daily and are causally associated with the number of cases, have been ignored in the model specification. Several authors have explicitly accounted for extra variation when using Poisson models [7,14,20].

The goal of this study was to develop a statistical model to serve as the basis for daily "bioterrorism alert" thresholds for the number of cases of ILI. To be operationally feasible, quantification of the threshold must be based on a few readily-obtainable pieces of information and the threshold must be adjustable to allow the Health Department to control the number of false alarms at a manageable level. The method developed can be classified as a regional surveillance model [21].

2. ILI DATA FOR ERIE COUNTY, NY

The data used to derive thresholds for event detection were obtained from the CUBRC)/Erie County Health Department data collection system for real-time monitoring of ILI incidence. Emergency services providers and selected emergency care centers throughout Erie County, NY reported data daily. We obtained daily counts of cases from August 1, 1999 through November 31, 2002. The dataset consisted of 1,217 observations (counts were missing for two days). Temporal patterns in the numbers reported are believed to be representative of those of the total numbers of such cases in the County. During this time period, no events occurred that would have pushed the mean number of cases above baseline. Reporting of these data was voluntary, and therefore subject to timing and variability problems [7]. Specifically, the decision to model daily counts (rather than, say, a weekly aggregate) was based on assumptions that reporting delays are negligible. Wagner [22] provided a detailed discussion of issues surrounding timely reporting.

Daily readings of temperature and precipitation at the Buffalo Niagara International Airport were obtained from the National Climactic Data Center Archives and were merged with case data. The resulting dataset was divided into training and validation subsets that included observations from August 1, 1999 through December 31, 2001 (884 days), and from December 1, 2001 through Nov 30, 2002 (365 days), respectively. December 2001 was included in both sets, for reasons discussed later.

***** (insert Figure 1 here) *****

3. MODELING AND THRESHOLD DEVELOPMENT

3.1. Simple Poisson Model Analysis

Let Y_t denote the number of cases reported on the t^{th} day. Suppose that the Y_t are independent identically distributed observations from a Poisson(μ) distribution. Then $P(Y = y) = \mu^y e^{-\mu} (y!)^{-1}$. Table 1 gives the observed frequency of days for counts between 1 and 32 inclusive. The maximum likelihood estimator of μ was 10.07, the sample mean number of cases. The expected numbers of days for each observed count under the Poisson distribution with $\mu = 10.07$ are given in the last column of table 1.

***** (insert Table 1 here) *****

A Chi-Square goodness of fit test resulted in a rejection of this distribution (p=0.00). The sample variance was 13.83, suggesting significant extra Poisson variation.

If one were to ignore the fact that the Poisson distribution does not fit the data then, denoting the acceptable false positive rate by p, the $100(1\text{-}p)^{th}$ percentile of the Poisson(10.07) would be used as the threshold number of cases to control the false positive alert rate at approximately $100p\%$. If p is set at 0.051, then the threshold would be 15 under the Poisson assumption. We see from table 1 that 125 days (10.3%) exceeded this threshold.

3.2. Negative Binomial Analyses

Agresti [23] recommended use of the Negative Binomial distribution when extra Poisson variation is observed. A Negative Binomial distribution with mean 10.07 and dispersion parameter $k = 27.02$ fits the observed responses very well (p=1.00). Under the best fitting Negative Binomial distribution, the 94.7% upper prediction limit was 16 cases per day. Results from this simple model derived from the training data and applied to the validation data will be used as a basis for comparison to the results from a more complex model presented below, which included weather variables and lagged numbers of cases.

3.3. Poisson Regression Model Analysis

Preliminary analyses of the data showed that month of the year and day of the week have significant effects on the number of cases. Furthermore, the number of cases was shown to be associated with weather. It seems likely, therefore, that the extra Poisson variation observed in the data was due at least in part to the influence of these variables on the mean. In that case, the counts are not "identically distributed". It also seems likely that recent observations would be correlated with the current day's observation, rendering the

independence assumption invalid. To address these issues, we performed a Poisson regression analysis with month, day of the week, temperature, precipitation, and lagged number of cases in the model.

Let Y_t be the number of ILI cases on a given day. We assume $Y_t \sim Poisson(\mu_t)$, where $\mu(x_t'\beta)$ is a monotone function of a linear combination of the predictor variables ($x_t'\beta$). Then we can write

$$Y_t = \mu(x_t'\beta) + \varepsilon_t,$$

where the vector of predictor variables includes Y_{t-1}. We specify $\mu(x_t'\beta)$ to be $\exp(x_t'\beta)$. That is, we assume the log link function $\log \mu_t = x_t'\beta$. We further assume that dependence of the Y_t observations over time is explained totally by the assumed dependence of the mean function on Y_{t-1}. That is, we assume that the ε_t are independent.

3.3.1. Model Selection

In the initial model specification, the coefficient vector, β, consisted of coefficients on 42 explanatory variables plus an intercept term. The explanatory variables included a one-day lagged incidence count, temperature and precipitation for the current and previous seven days, same-day interaction terms for temperature and precipitation, and indicator variables for month and day of the week.

The following variable selection strategies were employed on the training dataset:

1. A significance level of $\alpha = 0.05$ was used throughout the analysis.
2. Backwards Elimination Rules: In general, the variable with the largest p-value was dropped from the model. Interaction terms were considered first. If an interaction term was retained in a model, the associated main effects were also retained, regardless of their p-value.
3. An insignificant variable was retained if its removal caused another significant variable to become insignificant. This prevents suppression of effects due to multicollinearity.
4. Day of Week and Month indicator variables were retained throughout the model building process.

SAS version 9.1 Genmod Procedure was used to fit the models.

The backward selection process resulted in a final model, defined by the mean function

$$\mu(\eta_t) = \exp(x_t'\beta)$$

where

$$
\begin{aligned}
x_t'\beta = \beta_0 &+ \beta_L Y_{t-1} \\
&+ \sum_{i=0,1,3,4} \beta_{T(i)} \text{Temp}_i + \sum_{i=1,3,4,6,7} \beta_{P(i)} \text{Precip}_i \\
&+ \sum_{i=1,3,4} \beta_{TP(i)} \left(\text{Temp}_i \cdot \text{Precip}_i \right) \\
&+ \sum_{j=1}^{6} \beta_{D(j)} \left(\text{Week Day} \right)_j \\
&+ \sum_{j=1}^{11} \beta_{M(j)} \left(\text{Month} \right)_j
\end{aligned} \tag{1}
$$

Various diagnostic measures indicted a good fit for the final model. Compared with the deviance of the full model the backward model selection process did not compromise the goodness of fit (p=0.485). The Chi square and Standardized Deviance Residual QQPlots in figure 2 show that the final model fit very well.

***** (insert Figure 2 here) *****

Coefficient estimates from the final model are shown in table 2. The significance of P03, P04 and associated interaction terms corroborates the notions driving preliminary covariate selection. Variability of the flu incubation period and symptom severity may account for the range of precipitation lags (P01, P06 and P07).

***** (insert Table 2 here) *****

3.3.2. Derivation of Prediction Intervals

Consider the dependent variable $Y_t \sim Poisson\left(\mu\left(\eta_t\right)\right)$. Define η_t as the linear combination $x_t'\beta$. The model for the next observation Y_f is written as

$$Y_f = \mu\left(\eta_f\right) + \varepsilon_f$$

where ε_f ~ Centered Poisson$\left(\mu\left(\eta_f\right)\right)$ with $\text{var}\left(\varepsilon_f\right) = \mu\left(\eta_f\right)$. The associated prediction equation is

$$\hat{Y}_f = \mu\left(\hat{\eta}_f\right)$$

where the non-linear mean function under our model specification is

$$\mu\left(\eta_f\right) = \exp\left[x_f'\beta\right]. \tag{2}$$

By standard maximum likelihood theory, the maximum likelihood estimator $\hat{\beta}$ satisfies

$$\hat{\beta} \overset{app}{\sim} N\left(\beta, I^{-1}_{\beta,x'}\right)$$

when n is large, where $I_{\beta,x'}$ is the estimator of the information matrix, which is defined as

$$-E\left(\frac{\partial^2}{\partial\beta\partial\beta'}\log L(\beta \mid X)\right)$$

As the sample size increases, the information also increases, and the $\text{var}\left(\hat{\beta}\right)$ decreases. Derivation of a prediction limit for Y_f depends on the variance of the non-linear function

$$Y_f - \hat{Y}_f = \mu\left(\eta_f\right) - \mu\left(\hat{\eta}_f\right) + \varepsilon_f$$

The non-linear function $\mu\left(\hat{\eta}_f\right)$ can be approximated by a Taylor series expansion around $\hat{\beta} = \beta$. We have that

$$\begin{aligned} Y_f - \hat{Y}_f &= \mu\left(\eta_f\right) - \mu\left(\hat{\eta}_f\right) + \varepsilon_f \\ &\cong \mu\left(\eta_f\right) - \left[\mu\left(\eta_f\right) + \frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}}\right]\Bigg|_{\hat{\beta}=\beta}\left(\hat{\beta}-\beta\right) + \text{Remainder} + \varepsilon_f \\ &\cong -\frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}}\Bigg|_{\hat{\beta}=\beta}\left(\hat{\beta}-\beta\right) + O_p\left(1/n\right) + \varepsilon_f \\ &\cong -\frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}}\left(\hat{\beta}-\beta\right) + \varepsilon_f \end{aligned} \tag{3}$$

where the remainder term in the Taylor expansion is $O_p\left(1/n\right)$. It follows that

$$\begin{aligned} \text{var}\left(Y_f - \hat{Y}_f\right) &= \text{var}\left(\frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}}\left(\hat{\beta}-\beta\right) + \varepsilon_f\right) \\ &= \frac{\partial\mu\left(\hat{\eta}_f\right)'}{\partial\hat{\beta}} V_{\hat{\beta}} \frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}} + \sigma^2_{\varepsilon_f} \\ &\cong \frac{\partial\mu\left(\hat{\eta}_f\right)'}{\partial\hat{\beta}} I^{-1}_{\hat{\beta},x'} \frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}} + \sigma^2_{\varepsilon_f} \end{aligned} \tag{4}$$

because $E\left(\varepsilon_f\right) = 0$, $E\left(\hat{\beta}-\beta\right) = 0$, ε_f is independent of $\hat{\beta}$ and $V_{\hat{\beta},x'} \cong I^{-1}_{\beta,x'}$ is the variance/covariance matrix of $\hat{\beta}$. The goal is to calculate a $100(1-\alpha)\%$ upper prediction limit for Y_f with the form

$$\hat{Y}_f + C_{\alpha,f} \tag{5}$$

where $C_{\alpha,f}$ is a function of $\text{var}\left(Y_f - \hat{Y}_f\right)$. To that end, define $C_{\alpha,f}$ such that

$$\Pr\left(Y_f \le \hat{Y}_f + C_{\alpha,f}\right) = 1-\alpha \tag{6}$$

From equation (3), $Y_f - \hat{Y}_f$ is a function of an approximately normal term

$$-\frac{\partial\mu\left(\hat{\eta}_f\right)}{\partial\hat{\beta}}\left(\hat{\beta}-\beta\right)\overset{app}{\sim}N\left(0,\frac{\partial\mu\left(\eta_f\right)'}{\partial\beta}I^{-1}_{\hat{\beta},x'}\frac{\partial\mu\left(\eta_f\right)}{\partial\beta}\right) \quad (7)$$

and a centered Poisson term $\left(\varepsilon_f\right)$. We assume that the model appropriate for the past observations is applicable to the new observation. Maximum likelihood methods provide consistent estimators of all the required parameters. Given the maximum likelihood estimates, Monte Carlo simulation techniques can be used to estimate $C_{\alpha,f}$ for each new observation.

3.3.3. Algorithm for Estimating $C_{\alpha,f}$

The following algorithm uses Monte Carlo simulation to estimate $C_{\alpha,f}$ such that equation (6) is satisfied approximately. The strategy is to generate random observations from the two distributions represented in equation (3) given x_f, add the observations together, and chose the $1-\alpha$ quantile as $C_{\alpha,f}$. From equation (2) and equation (4), $\mu\left(\eta_f\right)$ and $\text{var}\left(\varepsilon_f\right)$ depend on the newly observed explanatory variables x'_f. As such, it is necessary to recalculate $\mu\left(\hat{\eta}_f\right)$ and the associated derivatives for each new observation. The algorithm for estimating $C_{\alpha,f}$ follows:

(1) Calculate $\hat{Y}_f$ from equation (2), using the estimate $\hat{\beta}$ calculated from the existing observations and x_f from the new observation. That is,

$$\hat{y}_f=\mu\left(\hat{\eta}_f\right)=\exp\left(x'_f\hat{\beta}\right);$$

(2) Again using equation (2), estimate $\frac{\partial}{\partial\beta}\mu\left(\eta_f\right)$ by substituting $\hat{\beta}$ for β. From

$$\frac{\partial}{\partial\beta}\mu\left(\eta_f\right)=\frac{\partial}{\partial\beta}\exp\left(x'_f\beta\right)$$

$$=\begin{bmatrix}x_{f1}\exp\left(x'_f\beta\right)\\x_{f2}\exp\left(x'_f\beta\right)\\\vdots\\x_{fk}\exp\left(x'_f\beta\right)\end{bmatrix}$$

we obtain

$$\frac{\partial}{\partial\beta}\mu\left(\hat{\eta}_f\right)\cong\begin{bmatrix} x_{f1}\exp\left(x_f'\hat{\beta}\right)\\ x_{f2}\exp\left(x_f'\hat{\beta}\right)\\ \vdots\\ x_{fk}\exp\left(x_f'\hat{\beta}\right)\end{bmatrix}\cong\hat{Y}_f\begin{bmatrix} x_{f1}\\ x_{f2}\\ \vdots\\ x_{fk}\end{bmatrix};$$

(3) Using the results from steps 1 and 2, estimate $\frac{\partial\mu\left(\eta_f\right)}{\partial\beta'}I^{-1}_{\hat{\beta},x'}\frac{\partial\mu\left(\eta_f\right)}{\partial\beta}$ in equation (4). The estimate of $I^{-1}_{\beta,x'}$ is provided by SAS Proc Genmod;

(4) Generate M random observation (denoted A_i) from a

$$N\left(0,\frac{\partial\mu\left(\eta_f\right)}{\partial\beta'}I^{-1}_{\hat{\beta},x'}\frac{\partial\mu\left(\eta_f\right)}{\partial\beta}\right)$$

distribution with covariance matrix taken to be the estimate in step (3);

(5) Generate M random observation (denoted B_i) from a $Poisson\left(\hat{Y}_f\right)$ distribution

(6) Calculate $A_i+B_i-\hat{Y}_f$, i=1, 2 … M, to obtain a random sample from the approximate distribution of $Y_f-\hat{Y}_f$.

(7) To satisfy equation (6), identify $C_{\alpha,f}$, the value of $A_i+B_i-\hat{Y}_f$ such that

$$\frac{1}{M}\sum_{i=1}^{M}I\left(A_i+B_i-\hat{Y}_f\le C_{\alpha,f}\right)=1-\alpha$$

(8) Then $\hat{Y}_f+C_{\alpha,f}$ is an approximate upper $100\left(1-\alpha\right)\%$ prediction limit for Y_f

4. PERFORMANCE OF COMPETING THRESHOLDS

Model coefficient estimators $\left(\hat{\beta}\right)$ and the associated covariance matrix were estimated for the final model in equation (1). The estimates were used to generate the predictions $\hat{Y}_f$ and $C_{\alpha,f}$ for each observation in both the training and validation datasets. M = 10,000 random samples per observation were generated for the Monte Carlo simulations. Prediction limits were then calculated for each observation. In the reporting that follows, the December 2001 has been included in the validation dataset, in order to avoid skewed comparisons of methods, in spite of the fact that it was also included in the calibration dataset.

Table 3 shows the performance of thresholds, derived as 95, 97, and 99% prediction limits. The 95% prediction limit identified 6.51% of the training set and 8.77% of the validation set as days with an unexpectedly high number of ILI cases. Possible

explanations of the fact that more than the expected percentages of the observations in the validation set exceeded the thresholds will be discussed in the next section.

***** (insert Table 3 here) *****

The numbers of days exceeding the prediction threshold were evenly spread across months and days of the week. (See table 4). This is a favorable indication of model performance.

***** (insert Table 4 here) *****

In building the Poisson regression model and validating the thresholds derived from it, we assumed that no unusual events occurred during the observation period. If this assumption is true, then the number of days that exceed the prediction threshold provides an indication of the false alarm rate. The validation results in table 3 suggest that the 95% prediction limit will result in about 32 false alarms during a year with no unusual incidents. The prediction limit level can be adjusted to accommodate the Health Department's tolerance for false alarms.

The number of false alarms generated by the modeled 95% prediction limit was about 6% below that for the simple Negative Binomial model, suggesting better performance of the Poisson regression model, and considerable cost savings to the County. The superiority of the regression model also is manifested by the more even distribution of false alarms between winter and summer months, in comparison to the negative binomial model.

5. CONCLUSIONS AND DISCUSSION

We conclude that the simple Poisson model with no covariates suffers from over dispersion; that the alternative Negative Binomial model fits well but produces a higher false alarm rate than the Poisson regression model; and, hence, that thresholds should be derived taking into consideration weather, season, day of the week, and any other variables that may be readily available in a timely fashion and are associated with ILI.

For a correctly specified model, prediction limits converge to corresponding tolerance limits as the sample size increases, but the sample prediction limits are expected to be wider than population tolerance limits in finite samples. Thus, it is surprising that the percentages of days that exceed prediction limits (see table 3) are higher than the nominal 1, 3, and 5% levels. The over dispersion evinced by this finding may be attributable to a variety of factors, including reporting errors, exclusion of other meaningful covariates, or a violation of our assumption that the ε_t are independent. The effects of seasonality and the infectious nature of the flu suggest some correlation among daily numbers of ILI

cases. We have included month of the year in our model to adjust for seasonal effects and Y_{t-1} to account for autocorrelation that should be expected when studying infectious illnesses. Kafadar [24] examined ratio of current (time series) counts to historic baselines when observations are correlated. Correlation induced an understated estimate of the ratio's variance, potentially causing the prediction thresholds to be too low. This is consistent with the results of table 3.

It is possible that additional lagged counts should be included in our model. Ignoring them would induce an auto-regressive error structure, in which case Y_{t-1} would be correlated with ε_t. This generally results in biased, inconsistent estimates of standard errors. The effect can be corrected by using instrumental variable estimation of the linear distributed lag model in the Generalized Linear Model framework. Under normality assumptions and with linear models, instrumental variable techniques are well documented [25]. An instrumental variable must be correlated with Y_{t-1}, uncorrelated with the error (ε_t) and act indirectly on the outcome Y_t through Y_{t-1}. Potential instrumental variables include prior weather variables. Unfortunately, more work is needed to develop instrumental variable methods for Generalized Linear distributed lag models.

Fortunately, one can heuristically assume that Y_{t-1} and ε_t are uncorrelated and use the methods of the current paper to derive thresholds, provided the false positive rate is estimated from a validation dataset. In the case of Erie County, for example, we would recommend the use of the 97% prediction limit in order to ensure a false alarm rate of about 5% (see table 3). Nevertheless, we would expect instrumental variable methods to be more efficient in the sense that, given the same sample size, a false alarm rate less than 5% would be expected of the threshold derived from an instrumental variable based 97% prediction limit. It is left for future research to develop the instrumental variables approach.

REFERENCES

1. Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models, *Statistics in Medicine* 1999; **18**(24):3463-3478. DOI: 10.1002/(SICI)1097-0258(19991230)18:24<3463::AID-SIM409>3.0.CO;2-I
2. Rath TM, Carreras M. Automated detection of influenza epidemics with Hidden Markov Models, *Lecture Notes in Computer Science* 2003; Volume 2810/2003, ISSN 0302-9743
3. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS), *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 2003; Vol 80, Supplement 1: i89-i96.
4. Rossi G, Lampugnani L, March M. An approximate CUSUM procedure for surveillance of health events, *Statistics in Medicine* 1999; 18:2111-2122
5. Rolfhamre P, Ekdahl K. An evaluation and comparison of three commonly used statistical models for automatic detection of outbreaks in epidemiological data of communicable diseases. *Epidemiol Infect*. 2006; **134**(4):863-71. DOI: 10.1017/S095026880500573X
6. Cowling BJ, Wong IOL, Ho LM, Riley S, Leung GM. Methods for monitoring influenza surveillance data, *International Journal of Epidemiology* 2006; **35**(5):1314-1321. DOI:10.1093/ije/dyl162.
7. Farrington CP, Andrews NJ. A statistical algorithm for the early detection of outbreaks of infections disease, *Journal of the Royal Statistical Society (series A)* 1996; **159**(3):547-563.
8. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Rep* 1963; 78: 494-506.
9. Parker RA. Analysis of surveillance data with Poisson regression: a case study, *Statistics in Medicine* 1989; 8:285-294
10. Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective analysis, *Epidemiol. Infect* 1999; 122:103–10. DOI:10.1017/S0950268898001939
11. Zeger SL. A regression model for time series of counts. *Biometrika* 1988; 75:621–29. DOI:10.1093/biomet/75.4.621
12. Watier L, Richardson S. Time series construction of an alert threshold with application of S. Bovismorbificans in France, *Statistics in Medicine* 1991; **10**(10):1493-1509.
13. Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. (2nd edn) Oxford University Press: Oxford, 2002.
14. Held L, Hohle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts, *Statistical Modeling* 2005; **5**(3):187-199. DOI: 10.1191/1471082X05st098oa.
15. Huntington E. The Control of Pneumonia and Influenza by the Weather, *Ecology* 1920;**1**(1):6–23. DOI: 10.1175/1520-0493(1920)48<501:TCOPAI>2.0.CO;2
16. Schulman JL, Kilbourne ED. Experimental transmission of influenza virus infection in mice. *Journal of Experimental Medicine* 1963; **118**(1):267-275.
17. Kilbourne ED. *Influenza.* Plenum Publishing: New York, 1987.
18. Beveridge WB. *Influenza: The Last Great Plague*, Prodist: New York, 1977.

19. Moser MR. An outbreak of influenza aboard a commercial airliner, *American Journal of Epidemiology* 1979; **110**(1):1-6.
20. Hakulinen T. Precision of incidence predictions based on Poisson distributed observations. *Statistics in Medicine* 1994; **13**(15):1513–1523.
21. Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, Pavlin JA, Gesteland PH, Treadwell T, Koski E, and others. Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience, *Journal of the American Medical Informatics Association* 2004; **11**(2):141. DOI 10.1197/jamia.M1356.
22. Wagner MM, Tsui FC, Espino JU, Dato VM, Sittig DF, Caruana RA, McGinnis LF, Deerfield DW, Druzdzel MJ, Fridsma DB. The Emerging Science of Very Early Detection of Disease Outbreaks, *J Public Health Management Practice* 2001; **7**(6):51–59.
23. Agresti, A. *Categorical Data Analysis* (2nd edn).Wiley: New Jersey, 2002. DOI 10.1002/0471249688.fmatter
24. Kafadar K, Stroup DF. Analysis of aberrations in public health surveillance data: estimating variance in correlated samples, *Statistics in Medicine* 1992; **11**(12): 1551-1568.
25. Fuller W. *Introduction to Statistical Time Series (2nd edn).* Wiley: New York, 1996.

Table 1: ILI Count Frequency Distribution

Daily Count	Observed Frequency	Expected Frequency	Contribution to Chi-Sq Statistic
1	2	0.52	4.24
2	8	2.61	11.15
3	10	8.75	0.18
4	45	22.04	23.91
5	52	44.40	1.30
6	86	74.54	1.76
7	98	107.25	0.80
8	128	135.04	0.37
9	136	151.13	1.51
10	135	152.22	1.95
11	124	139.38	1.70
12	101	116.99	2.19
13	91	90.64	0.00
14	76	65.21	1.78
15	40	43.79	0.33
16	24	27.57	0.46
17	19	16.33	0.44
18	13	9.14	1.63
19	10	4.85	5.48
20	6	2.44	5.19
21	7	1.17	29.04
23	3	0.23	32.59
24	2	0.10	36.72
32	1	0.65	0.19

Table 2
Coefficient Estimates of the Final Fitted Poisson Regression Model

Parameter	Description	Estimate	StdErr	ProbChiSq
Intercept	Intercept	1.986	0.081	0.000
cnt01	Count Lag(1)	0.009	0.003	0.001
T00	Temperature Lag(0)	0.006	0.002	0.001
T01	Temperature Lag(1)	0.000	0.002	0.938
T03	Temperature Lag(3)	0.000	0.002	0.841
T04	Temperature Lag(4)	-0.001	0.002	0.676
P01	Precipitation Lag(1)	0.389	0.141	0.006
P03	Precipitation Lag(3)	0.380	0.145	0.009
P04	Precipitation Lag(4)	0.472	0.147	0.001
P06	Precipitation Lag(6)	0.097	0.046	0.036
P07	Precipitation Lag(7)	0.104	0.045	0.022
TP01	Interaction Lag(1)	-0.007	0.003	0.018
TP03	Interaction Lag(3)	-0.006	0.003	0.037
TP04	Interaction Lag(4)	-0.007	0.003	0.011
M01	January	0.019	0.053	0.726
M02	February	-0.020	0.053	0.714
M03	March	-0.034	0.051	0.511
M04	April	-0.109	0.057	0.054
M05	May	-0.298	0.071	0.000
M06	June	-0.277	0.080	0.001
M07	July	-0.212	0.082	0.010
M08	August	-0.228	0.082	0.005
M09	September	-0.220	0.073	0.003
M10	October	-0.170	0.059	0.004
M11	November	-0.143	0.051	0.005
D1	Sunday	0.030	0.041	0.457
D2	Monday	0.084	0.040	0.036
D3	Tuesday	0.010	0.041	0.806
D4	Wednesday	0.042	0.041	0.294
D5	Thursday	0.017	0.041	0.677
D6	Friday	-0.009	0.041	0.833
Scale	Overdispersion	1.000		

Table 3

Percentage of Days Exceeding Simulation-derived Thresholds[‡]

Prediction Limit	Data Set	
	Training	Validation
95%	6.51%	8.77%
97%	3.08%	4.11%
99%	1.89%	2.74%

[‡] December 2001 data are included in both the training and validation sets

Table 4

Number of False Alarms by Month and Day of Week

Month	Total Days	Days Over Threshold	
		NegBin	Model
Jan	31	7	4
Feb	28	3	3
Mar	31	3	2
Apr	30	0	0
May	31	2	4
Jun	30	2	2
Jul	31	1	2
Aug	31	2	2
Sep	30	1	2
Oct	31	5	6
Nov	30	1	2
Dec	31	7	3
Total	**365**	**34**	**32**

Day of Week	Total Days	Days Over Threshold	
		NegBin	Model
Sunday	52	4	3
Monday	52	3	2
Tuesday	52	6	6
Wednesday	52	6	6
Thursday	52	6	6
Friday	52	6	7
Saturday	53	3	2
Total	**365**	**34**	**32**

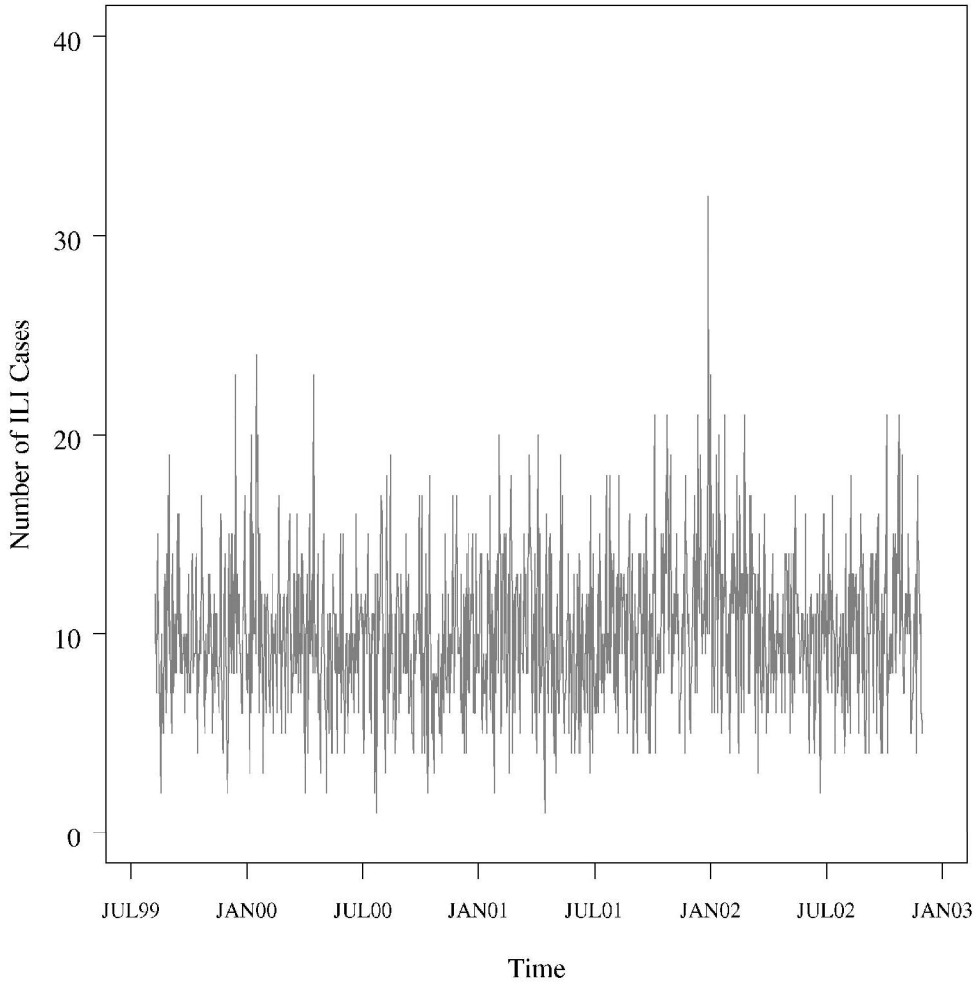

Figure 1: The daily number of ILI cases reported by emergency services personnel in Erie County NY between August 1, 1999 and November 31, 2002.
101x101mm (600 x 600 DPI)

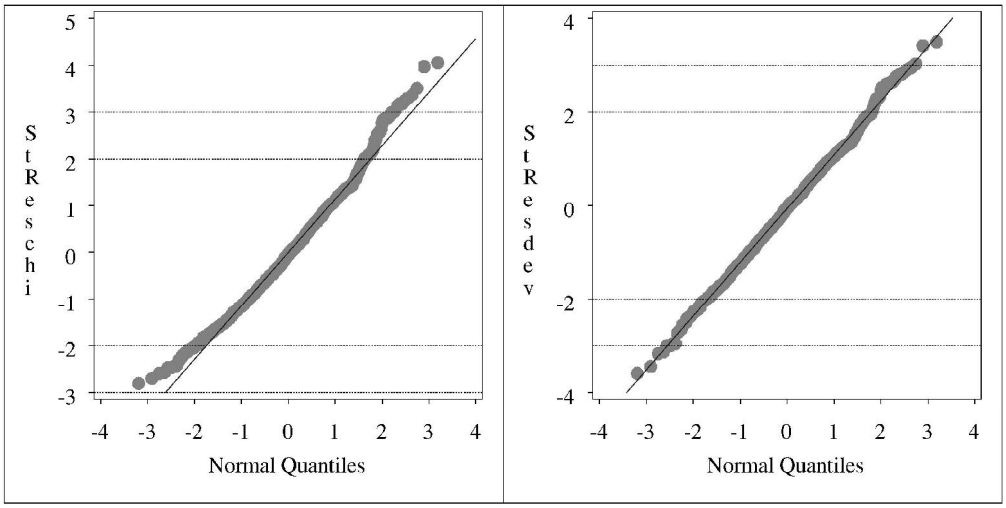

Figure 2: Chi square and standardized deviance residuals demonstrate the goodness of fit of the final Poisson regression model to the observed daily flu counts.
127x63mm (600 x 600 DPI)