M. Baron. Bayes and asymptotically pointwise optimal stopping rules for the detection of influenza epidemics. In *C. Gatsonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler and I. Verdinelli, Eds., Case Studies in Bayesian Statistics*, vol. 6, pages 153–163, Springer-Verlag, New York, 2002.

Bayes and asymptotically pointwise optimal stopping rules for the detection of influenza epidemics

Michael I. Barón

ABSTRACT Whereas it is customary to announce epidemics when influenza mortality exceeds the epidemic threshold, one can often detect the beginning of epidemics earlier, by solving a suitable change-point problem. We propose a hierarchical Bayesian change-point model for influenza epidemics. Prior probabilities of a change point depend on (random) factors that affect the spread of influenza. Theory of optimal stopping is used to obtain Bayes stopping rules for the detection of epidemic trends under the loss functions penalizing for delays and false alarms. The Bayes solution involves rather complicated computation of the corresponding payoff function. Alternatively, asymptotically pointwise optimal stopping rules can be computed easily and under weaker assumptions. Both methods are applied to the 1996–2001 influenza mortality data published by CDC.

Key words and phrases: asymptotically pointwise optimal, Bayes sequential, change point, epidemic threshold, influenza mortality, optimal stopping, payoff function.

1 Introduction

Influenza causes more morbidity and mortality in the United States than AIDS. Each year it accounts for 10,000 to 40,000 fatalities, nearly 200,000 hospitalizations, and about 70 million work-loss days, with the associated annual cost of $12 billion. The rate of influenza diagnoses and influenza related hospitalizations rapidly increases during epidemic periods (Bridges et al, 2000; Neuzil et al., 1999).

Influenza epidemics are determined when influenza mortality, the proportion of deaths attributed to influenza, exceeds the *epidemic threshold* (Figure 1). As seen on the Figure, several weeks before the beginning of epidemics are marked by unusually high increase of mortality. The goal of this article is to determine such *pre-epidemic trend* in order to detect the beginning of epidemics even before the epidemic threshold is exceeded.

Recent studies (Medina et al., 1997; Peters et al., 2000) confirmed strong association between influenza epidemics and weather conditions, air pollu-

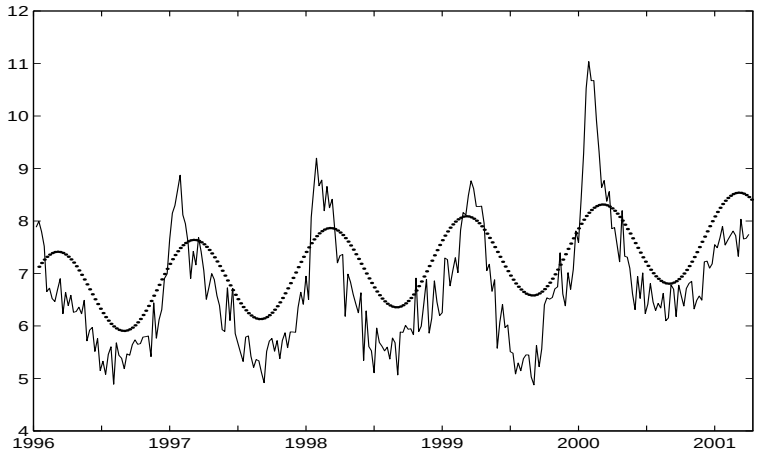

FIGURE 1. Influenza mortality: percentage of deaths attributed to influenza-like illnesses, reported by CDC

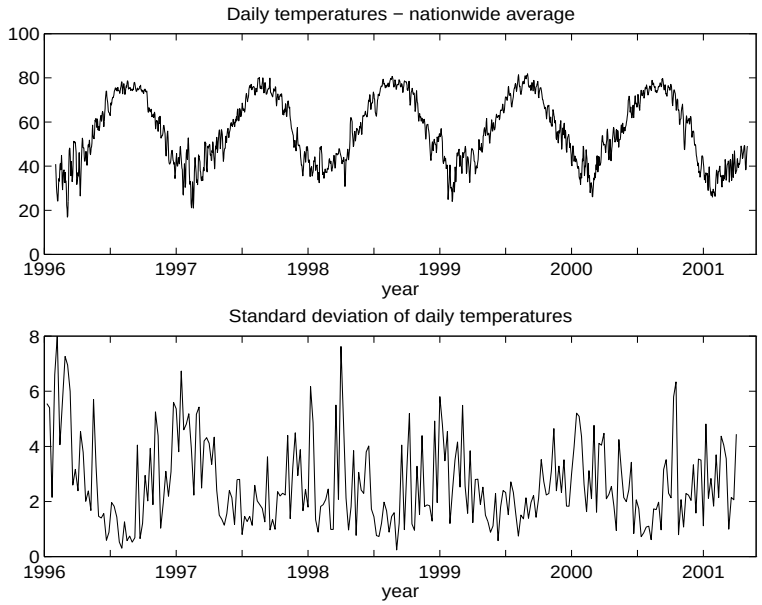

FIGURE 2. Average daily temperatures in U.S. and weekly standard deviations of temperatures

tion, pollen, ozone level, and other factors. For example, the likelihood of an influenza epidemic increases during a passage of a cold front followed by a high pressure system. At the same time, the influenza virus dies more rapidly with high humidity and vigorous air movement (White and Hertz-Picciotto, 1985).

In this article, we propose and apply a method of early detection of the beginning of epidemics based on weather conditions and weekly influenza mortality, reported by The Centers for Disease Control and Prevention (CDC).

Every pre-epidemic period is characterized by rapidly increasing mortality in excess of the general trend (Figure 1). Detrended mortality Z_t has a positive drift for few weeks before exceeding the epidemic threshold, while being stationary between epidemics. This suggests a *change-point model* where the distribution of increments $X_t = \nabla Z_t$ changes from zero-centered to positive-centered. The problem of detecting pre-epidemic trends becomes a *sequential change-point problem*.

Frequent changes of temperature increase the probability of influenza. Figures 1 and 2 show the association between influenza mortality and instability of weather, quantified by the weekly standard deviation of average daily temperatures V_t. This justifies the prior distribution π_t of the change point parameter ν determined by the observed and forecasted values of V_t. In reality, an accurate forecast of V_t is available only for the next few days. Therefore, it is convenient to express π_t in terms of the *discrete hazard function* of ν,

$$\phi_t = \mathbf{P}\left\{\nu = t \mid \nu \geq t\right\},$$

and to assume that only $\{\phi_t,\ t \leq N\}$ are known at time N. It should also be noticed that V_t represents a stochastic process, which brings the hyper-prior distribution into consideration.

The corresponding hierarchical stochastic model is described in detail in Section 2. It is used in Section 3 to obtain the Bayes stopping rule for the detection of pre-epidemic trends. The method is based on the theory of optimal stopping (Shiryaev, 1978; Chow et al., 1991) and involves a rather complicated *payoff function*. Computation of the payoff function can only be done by an iterative numeric algorithm. Therefore, it seems feasible only for relatively simple prior distributions and loss functions.

To overcome these difficulties, we propose to use *asymptotically pointwise optimal (APO)* stopping rules (Bickel and Yahav, 1967, 1968) which represent sensible approximation of Bayes sequential procedures. Our definition of an APO rule is slightly different from that of Bickel and Yahav, because in change-point problems the cost function should penalize for after-change observations only, hence it is not linear. Nevertheless, a result similar to Bickel and Yahav (1968) can be obtained. Contrary to the Bayes procedure, the APO stopping rule has a closed form expression that is ready for application without an extensive numerical routine. This also

allows more complicated and rich models for the prior distribution of the change point.

We apply the proposed methods to nationwide data on weather and influenza mortality. More accurate results can be obtained locally, applying the same scheme to regional data.

2 Model

The observed (reported) time series Y_t represents the proportion of deaths attributed to influenza-like illnesses during week t. Clearly (see Figure 1) Y_t is marked by a general trend m_t and a seasonal component s_t,

$$Y_t = m_t + s_t + Z_t. \tag{2.1}$$

The error term Z_t may be viewed as deviation from the "normal" mortality. Pre-epidemic patterns can be recognized by an unusually fast increase of mortality beyond the general trend. Therefore, we consider the *differenced* series

$$X_t = \nabla Z_t = Z_t - Z_{t-1}. \tag{2.2}$$

During the "control" phase, i.e., between the epidemics, the process Z_t is stationary, and $\{X_t, t \le \nu\}$ is assumed to be a gaussian white noise with mean 0 and variance σ^2. A pre-epidemic trend starts during week ν, after which the increments of Z_t have positive expectations. This is modeled by a normal distribution of $X_t, t > \nu$, with mean $\mu > 0$ and variance τ^2. The change point ν represents the beginning of a pre-epidemic trend. It is the parameter of interest.

Assumption of normality agrees with the historical data, but it is not crucial. For the fast detection of epidemics, one needs a parametric model, because any nonparametric algorithm incurs a rather long delay after the change occurs (Baron, 2000).

The prior distribution of ν is expressed by the discrete hazard function

$$\phi_t = \mathbf{P}\left\{\nu = t \mid \nu \ge t\right\}. \tag{2.3}$$

At time N, only $\{\phi_t,\ t \le N\}$ are known. That is, during any week, we know only the probability of a change occurring before the end of the week, if it has not already occurred. These probabilities, ϕ_t, are determined by weather conditions. In the most simple scenario, we consider *favorable* and *unfavorable* weeks for the beginning of influenza epidemics. Favorable weeks are essentially the weeks of unstable weather. The probability of a change during a favorable week is higher than during an unfavorable week. Thus,

$$\phi_t = \begin{cases} p_0 & \text{on unfavorable weeks} \\ p_1 & \text{on favorable weeks,}\ \ p_1 > p_0 \end{cases} \tag{2.4}$$

A richer model is considered in Section 4.

Favorable and unfavorable weeks alternate according to a Markov chain. Hence, the prior probabilities $\{\phi_t\}$ have a *hyper-prior distribution* that is also a Markov chain, with a transition probability matrix

$$\mathcal{P} = \begin{pmatrix} 1-q_0 & q_0 \\ q_1 & 1-q_1 \end{pmatrix}. \tag{2.5}$$

Unfavorable weeks are followed by favorable weeks with probability q_0, and favorable weeks are followed by unfavorable weeks with probability q_1. Parameters $\mu, \sigma, \tau, p_0, p_1, q_0$, and q_1 are estimated from the historical data.

3 Bayes stopping rules

Our goal is to estimate the change-point parameter ν sequentially. Sequential nature of the problem arises from the need to detect influenza epidemics "as soon as possible". A suitable loss function should include the delay term $(T-\nu)^+$, T being the stopping time, penalizing for every week after the beginning of an epidemic trend when the algorithm fails to detect the change point. Also, the loss function should penalize for false alarms, when $T < \nu$. Without this, the optimal decision that minimizes the mean delay is to report a change point every week. This justifies the risk function

$$R(T, \{\phi_t\}) = \lambda \mathbf{E}(T-\nu)^+ + \mathbf{P}\{T < \nu\}, \tag{3.1}$$

combining the *mean delay* $\mathbf{E}(T-\nu)^+$ and the probability of false alarms $\mathbf{P}\{T < \nu\}$. A stopping rule T^* is *Bayes* if it minimizes $R(T, \{\phi_t\})$ over all stopping rules T.

According to the theory of optimal stopping, the Bayes stopping rule can be obtained as

$$T^* = \inf\{t \mid s(U_t) = \zeta(U_t)\}, \tag{3.2}$$

if there exist functions $\eta(\cdot)$ and $\zeta(\cdot)$ and a Markov sequence $\{U_n\}_{n=0}^{\infty}$, such that

$$R(T, \{\phi_t\}) = E\left\{\sum_{t<T} \eta(U_t) - \zeta(U_T)\right\} \tag{3.3}$$

(Shiryaev, 1978, Theorem 2.23). Here $s(u) = \sup_T \{-R(T, \{\phi_t\}) \mid U_0 = u\}$ is the *payoff function*, characterized as the solution of equation

$$s(u) = \max\{\zeta(u),\ \mathbf{E}\{s(U_{t+1}) \mid U_t = u\} - \eta(u)\}. \tag{3.4}$$

For the model described in Section 2, condition (3.3) is satisfied by functions $\eta(\pi, \cdot) = \lambda\pi$, $\zeta(\pi, \cdot) = \pi - 1$, and a homogeneous Markov sequence

$U_t = (\Pi_t, \phi_{t+1})$, where

$$\begin{aligned}\Pi_t &= \mathbf{P}\left\{\nu \le t \,\middle|\, \begin{array}{l} X_1,\ldots,X_t \\ \phi_0,\ldots,\phi_t \end{array}\right\} \\ &= \frac{\sum_{k\le t}\phi_k \prod_{j<k}(1-\phi_j)\prod_{j>k}\rho(X_j)}{\sum_{k<t}\phi_k\prod_{j<k}(1-\phi_j)\prod_{j>k}\rho(X_j)+1-\prod_{j<t}(1-\phi_j)}, \end{aligned} \tag{3.5}$$

and ρ is the likelihood ratio,

$$\rho(x) = \frac{f(x \mid 0,\sigma)}{f(x \mid \mu,\tau)} = \frac{\sigma}{\tau}\exp\left\{-\frac{1}{2}\left[\left(\frac{X_i-\mu}{\tau}\right)^2 - \left(\frac{X_i}{\sigma}\right)^2\right]\right\}.$$

Hence, the Bayes stopping rule has the form (3.2).

It remains to compute the payoff function $s(\cdot)$. Being the solution of the fixed-point functional equation (3.4), it can be computed as the limit

$$s(u) = \lim_{N\to\infty} \mathcal{Q}^N \zeta(u), \tag{3.6}$$

where $\mathcal{Q}$ is an operator defined as

$$\mathcal{Q}w(u) = \max\left\{w(u),\ \mathbf{E}\left\{w(U_{t+1}) \mid U_t = u\right\} - \eta(u)\right\}. \tag{3.7}$$

For the model (2.1)-(2.5), this is implemented as follows (Baron, 2001). Define an operator $\mathcal{T}$,

$$\mathcal{T}w(\pi,\phi) = \mathbf{E}\left\{w(U_{t+1}) \mid U_t = (\pi,\phi)\right\}, \tag{3.8}$$

independent of t because of the homogeneity of $\{U_t\}$. For $\phi \in \{p_0, p_1\}$ and any function w, one can compute $\mathcal{T}w(\pi,\phi)$ as

$$\begin{aligned}\mathcal{T}w(\pi,p_0) &= \mathbf{E}\left\{(1-q_0)w\left(\frac{\pi\rho(X)+(1-\pi)p_0}{\pi\rho(X)+1-\pi}, p_0\right)\right. \\ &\qquad \left. + q_0 w\left(\frac{\pi\rho(X)+(1-\pi)p_0}{\pi\rho(X)+1-\pi}, p_1\right)\right\}(\pi\rho(X)+1-\pi), \\ \mathcal{T}w(\pi,p_1) &= \mathbf{E}\left\{(1-q_1)w\left(\frac{\pi\rho(X)+(1-\pi)p_1}{\pi\rho(X)+1-\pi}, p_1\right)\right. \\ &\qquad \left. + q_1 w\left(\frac{\pi\rho(X)+(1-\pi)p_1}{\pi\rho(X)+1-\pi}, p_0\right)\right\}(\pi\rho(X)+1-\pi), \end{aligned} \tag{3.9}$$

where all the expectations are taken with respect to the normal$(0,\sigma^2)$ distribution of X. Based on (3.9), compute the operator $\mathcal{Q}$,

$$\mathcal{Q}w(u) = \max\left\{w(u), \mathcal{T}w(\pi,\phi) - \eta(u)\right\},$$

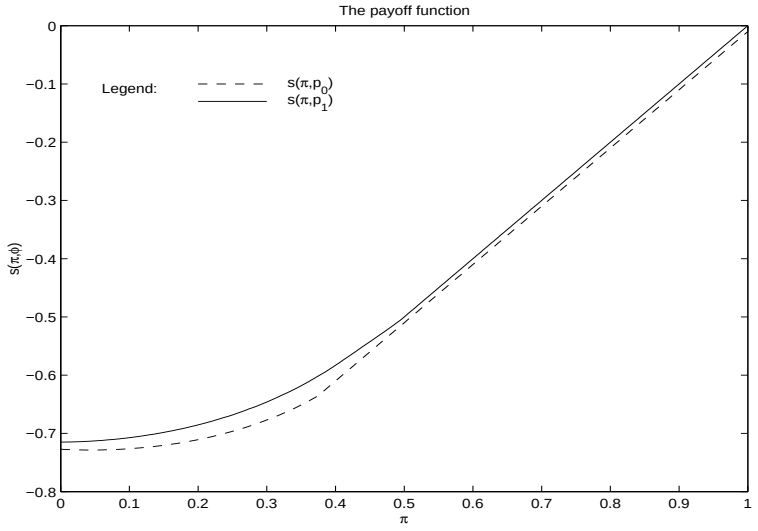

FIGURE 3. Payoff function $s(\pi, \phi)$ for the influenza mortality data

applying it to functions $\zeta(\pi, \phi)$, $\mathcal{Q}\zeta(\pi, \phi)$, $\mathcal{Q}\mathcal{Q}\zeta(\pi, \phi) = Q^2\zeta(\pi, \phi)$, etc., successively. A large number of iterations results in a sensible approximation of $s(u) = \lim_{N\to\infty} Q^N\zeta(u)$.

Then, according to (3.2), the Bayes stopping time equals

$$T^* = \inf\{t \geq 0 \mid \Pi_t \geq \pi^*\},$$

where the critical probability $\pi^*(\phi)$ is computed from the payoff function for each value of ϕ,

$$\pi^*(p_i) = \min\{\pi \mid s(\pi, p_i) = \zeta(\pi, p_i)\},\ i = 0, 1.$$

Application of this scheme to the CDC influenza mortality data resulted in a payoff function $s(\pi, \phi)$ depicted on Figure 3.

In this study, we used $\lambda = 0.03$ and defined a "favorable" week to be a week during the influenza season (October through March), when the standard deviation of the average daily temperature across U.S. exceeds 4 degrees Fahrenheit ($V_t > 4$). Estimates of the nuisance parameters were: $\mu = 0.0654$, $\sigma = 0.4278$, $\tau = 0.4893$, $\phi_0 = 0.01$, $\phi_1 = 0.04$, $q_0 = 0.2286$, $q_1 = 0.75$. Based on this payoff function, the critical probability equals, $\pi^*(0.01) = 0.375$ and $\pi^*(0.04) = 0.500$. Change points representing the beginning of influenza epidemics were detected on 12/28/1996, 01/10/1998, 10/24/1998, and 01/01/2000.

4 Asymptotically pointwise optimal stopping rules

The main disadvantage of the scheme described in the previous section is a rather complicated computation of the payoff function. Inevitable iterations imply that (3.9) can only be computed for finitely many values of ϕ. This was the main reason of choosing a simple model (2.4)-(2.5), with only two states of the Markov chain. Besides, even in the simplest case, the payoff function does not have a closed-form expression, and it can only be computed approximately.

Alternatively, the computation of *asymptotically pointwise optimal rules* (APO) for this problem is relatively easy. Bickel and Yahav (1967, 1968) define a stopping rule $\tilde{T} = \tilde{T}(c)$ to be APO under a loss function $L(\delta, \theta)$ and a linear cost function $c(n) = cn$ (more generally, $c(n) = cK(n)$) if

$$\limsup_{c \downarrow 0} \frac{\mathbf{E}\left\{L(\delta(\tilde{T}; X_1, \ldots, X_{\tilde{T}}), \theta) + c\tilde{T}\right\}}{\mathbf{E}\left\{L(\delta(T; X_1, \ldots, X_T), \theta) + cT\right\}} \leq 1 \tag{4.1}$$

for any stopping rule $T = T(c)$. Here δ is the Bayes terminal decision rule and θ is the unknown parameter. They also prove that (in a simplified formulation of Ghosh et al., 1997, Theorem 5.4.1) if

$$n^{\beta}\mathbf{E}\left\{L(\delta(n, X_1, \ldots, X_n), \theta) \mid X_1, \ldots, X_n\right\} \tag{4.2}$$

converges, as $n \uparrow \infty$, to a positive random variable a.s., then

$$\tilde{T} = \min\left\{n \mid n^{-1}\mathbf{E}\left\{L(\delta, \theta) + c\tilde{T}\right\} \leq c/\beta\right\} \tag{4.3}$$

is an APO stopping rule.

We notice that the risk function

$$\mathbf{E}\left\{L(\delta, \theta) + c\tilde{T}\right\} \tag{4.4}$$

is inappropriate for our change-point problem. As in (3.1), the mean delay term replaces the cost $c\tilde{T}$ in (4.4). However, under the risk (3.1), there is no convergence of (4.2) to a positive random variable.

Instead, we use the risk function

$$R_1(T, \nu) = \lambda\mathbf{E}(T - \nu)^+ - \log^{-1}\mathbf{P}\left\{T < \nu\right\} \tag{4.5}$$

that also penalizes for delays and false alarms. Similarly to (4.2) and (4.3), it can be shown that

$$n\mathbf{E}\left(-\log^{-1}\mathbf{P}\left\{n < \nu \mid X_1, \ldots, X_n\right\}\right) \tag{4.6}$$

has a positive limit a.s., and the stopping rule

$$\tilde{T} = \inf\left\{t : -t\log S_t \geq \frac{1}{\lambda}\right\} \tag{4.7}$$

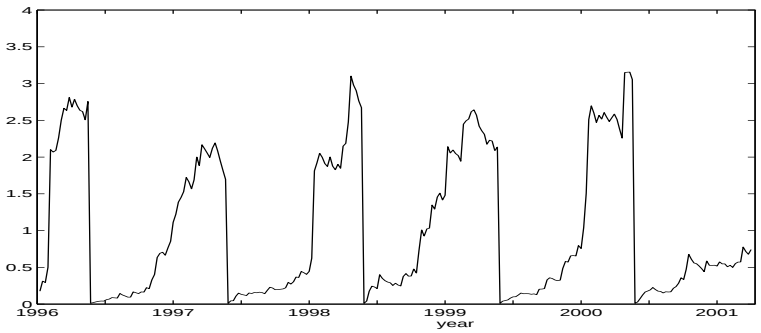

FIGURE 4. Negative log-posterior survival function $S_t(X_1, \ldots, X_t)$.

is APO. Here $S_t = S_t(X_1, \ldots, X_t) = \mathbf{P}\{\nu > t \mid X_1, \ldots, X_t\} = 1 - \Pi_t$ is the posterior survival function of the change point.

Computation of $\tilde{T}$ is straightforward. Posterior probabilities Π_t can be computed by (3.5). Results (4.6) and (4.7) are valid for any prior probabilities $\{\phi_t\}$, if there exists a limit, $\lim_{n\to\infty} t^{-1} \log \mathbf{P}\{\nu > t\} \in [-\infty, 0]$. This condition is satisfied by the prior defined by (2.4)-(2.5), but one can use considerably more complex prior distributions.

Extending the model with "favorable" and "unfavorable" days to a continuous case, we choose a discrete hazard function ϕ_t to be proportional to the weekly variance of daily average temperatures, $\phi_t \propto V_t^2$. The resulting statistic $(-\log S_t)$ is depicted on Figure 4.

Using this algorithm, epidemic trends were detected on 01/25/1997, 01/10/1998, 11/07/1998, and 01/08/2000, which constitutes an average of 3.25 weeks before influenza mortality exceeded the epidemic threshold. Figure 4 clearly shows low $(-\log S_t)$ for the 2000-2001 influenza season, which was the only season during the last 5 years when the epidemic threshold was not exceeded.

5 Conclusions and extensions

The proposed schemes are designed to obtain Bayes and APO stopping rules for the detection of epidemic or pre-epidemic trends, based on the reported influenza mortality and daily temperatures. The Bayes rule solves the problem of optimal stopping, whereas the APO rule represents an attractive approximation of the Bayes rule, for the weaker assumptions and simplicity of implementation. Both rules essentially solve the corresponding sequential change-point detection problem.

Both methods can be applied to other infectious diseases that can spread and cause epidemics, e.g. malaria (Wolfson, 2001). Other significant factors, air pollution, pollen, humidity, in addition to weather instability, can be

used to determine the prior distribution of a change point, "favorable" weeks, etc. Finally, one can use multivariate data, including, in addition to influenza mortality, sales of related grocery items (Goldenberg et al., 2001), the number of diagnosed patients (Baron, 2001), etc.

References

M. Baron (2000). Nonparametric adaptive change-point estimation and on-line detection. *Sequential Analysis*, 19:1–23.

M. Baron (2001). Bayes stopping rules in a change-point model with a random hazard rate. *Sequential Analysis*, 20 (4), to appear.

P. J. Bickel and J. A. Yahav (1967). Asymptotically pointwise optimal procedures in sequential analysis. In *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, volume I: Statistics, pages 401–413, Univ. California Press, Berkeley, Calif.

P. J. Bickel and J. A. Yahav (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Stat.*, 39:442–456.

C. B. Bridges, A. G. Winquist, K. Fukuda, N. J. Cox, J. A. Singleton, R. A. Strikas (2000). Prevention and Control of Influenza. *Morbidity and Mortality Weekly Reports, Centers for Disease Control and Prevention*, 49:1–38.

Y. S. Chow, H. Robbins and D. Siegmund (1991). *Great Expectations: The Theory of Optimal Stopping*. Dover Publications, New York.

M. Ghosh, N. Mukhopadhyay and P. K. Sen (1997). *Sequential Estimation*. Wiley, New York.

S. Medina, A. Le Tertre, P. Quénel, Y. Le Moullec, P. Lameloise, J. C. Guzzo, B. Festy, R. Ferry, W. Dab (1997). Air pollution and doctors' house calls: Results from the ERPURS system for monitoring the effects of air pollution on public health in Greater Paris, France, 1991-1995. *Environmental Research*, 75:73–84.

K. M. Neuzil, G. W. Reed, E. F. Mitchel, M. R. Griffin (1999). Influenza-associated morbidity and mortality in young and middle-aged women. *J. Amer. Medical Assoc.*, 281:901–907.

A. Peters, J. Skorkovsky, F. Kotesovec, J. Brynda, C. Spix, H. E. Wichmann, and J. Heinrich (2000). Associations between mortality and air pollution in Central Europe. *Environmental Health Perspectives*, 108:283–287.

A. N. Shiryaev (1978). *Optimal Stopping Rules*. Springer-Verlag, New York.

A. Goldenberg, G. Shmueli, and R. Caruana (2001). Using grocery sales data for the detection of bio-terrorist attacks. *Preprint.*

M. R. White and I. Hertz-Picciotto (1985). Human health: Analysis of climate related to health. In *Characterization of information requirements for studies of CO2 effects: Water resources, agriculture, fisheries, forests and human health, ed. M.R. White*, pages 172–206, Washington, D.C.: Department of Energy.

L. Wolfson (2001). Applicability of sequential change-point detection methods in the study of malaria in developing countries. *Private communication.*