A Test for Constant Fatality Rate of an Emerging Epidemic: With Applications to Severe Acute Respiratory Syndrome in Hong Kong and Beijing

K. F. Lam,[1,*] J. V. Deshpande,[2] E. H. Y. Lau,[1] U. V. Naik-Nimbalkar,[2] P. S. F. Yip,[1] and Ying Xu[1]

[1]Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong
[2]Department of Statistics and Centre for Advanced Studies, University of Pune, Pune 411007, India
*email: hrntlkf@hku.hk

SUMMARY. The etiology, pathogenesis, and prognosis for a newly emerging disease are generally unknown to clinicians. Effective interventions and treatments at the earliest possible times are warranted to suppress the fatality of the disease to a minimum, and inappropriate treatments should be abolished. In this situation, the ability to extract most information out of the data available is critical so that important decisions can be made. Ineffectiveness of the treatment can be reflected by a constant fatality over time while effective treatment normally leads to a decreasing fatality rate. A statistical test for constant fatality over time is proposed in this article. The proposed statistic is shown to converge to a Brownian motion asymptotically under the null hypothesis. With the special features of the Brownian motion, we are able to analyze the first passage time distribution based on a sequential tests approach. This allows the null hypothesis of constant fatality rate to be rejected at the earliest possible time when adequate statistical evidence accumulates. Simulation studies show that the performance of the proposed test is good and it is extremely sensitive in picking up decreasing fatality rate. The proposed test is applied to the severe acute respiratory syndrome data in Hong Kong and Beijing.

KEY WORDS: Brownian motion; Infectious disease; Real-time fatality rate; Severe acute respiratory syndrome; Zero-mean martingale.

1. Introduction

One of the key statistics for monitoring a newly emerging disease is the fatality rate. Traditionally, the fatality rate is assumed to be constant throughout the epidemic, and is generally estimated as the ratio of the cumulative number of deaths to the cumulative number of cases. It is well known that this estimator is insensitive to changes in fatality over the course of the epidemic (Last, 1995). Studies have shown that this traditional estimator has underestimated the fatality of the severe acute respiratory syndrome (SARS) considerably especially in the early part of the epidemic (Yip, Lam, et al., 2005; Yip, Lau, et al., 2005). The consequence of this underestimation could be very serious because the implementation of suitable preventive measures at the early stage of an outbreak may heavily depend on the estimated fatality rate. Moreover, the fatality rate of a newly emerging epidemic may vary over time possibly due to improvement in clinical treatments and hospital management. Sometimes, changes in weather or environmental conditions over the course of the epidemic may also have an effect on the fatality. It is important to keep track of the fatality rate over the course of the epidemic so that the effectiveness of any implemented health policy or clinical treatment can be assessed and evaluated. The estimator for the time-varying fatality rate via a counting process and kernel function was proposed in Yip, Lam, et al. (2005). An estimator derived from the

chain multinomial model was considered by Yip, Lau, et al. (2005).

In this article, we propose a test for testing the null hypothesis of constant fatality rate against the alternative hypothesis of decreasing fatality rate. Due to the urgency in an outbreak of an epidemic, effective interventions and treatments at the earliest possible time are warranted to suppress the fatality of the disease to a minimum. Hence, a desirable property of the test is the ability to reject the null hypothesis as soon as we have enough evidence of a decreasing fatality as information accumulates over time. This property is particularly important in practice when combating a newly emerging disease. For example, the etiology, pathogenesis, clinical features, clinical management, or prognosis for the SARS epidemic were almost unknown to clinicians. Therefore, appropriate and effective interventions should be undertaken at the earliest possible time during the outbreak or else any delayed effort will result in less impact in controlling the epidemic. In this situation, the ability to extract the most information out of the data available is critical for making a focused and effective decision. Therefore, it is desirable if we can reject the null hypothesis of constant fatality rate at the earliest possible time during the outbreak of the epidemic.

A constant fatality rate is, very likely, a result of the ineffectiveness of the current health policy or treatment in combating the disease or virus. In the SARS epidemic in 2003,

it was suggested by the World Health Organization (WHO) that simple methods for calculating case-fatality ratios from aggregate data will not give reliable estimates during the course of an epidemic (WHO, 2003). This statement is not entirely true. In the case of a constant fatality rate, a simple estimator based on the ratio of the cumulative number of deaths to the sum of the cumulative numbers of deaths and recoveries is shown to be its maximum likelihood estimator (MLE). This estimator is demonstrated to be more efficient in Section 3 and is more preferred if the assumption of constant fatality is valid. If the fatality rate is known to vary over time, a real-time fatality rate estimator should be used (Yip, Lam, et al., 2005; Yip, Lau, et al., 2005), which was shown to be able to capture the changes and to monitor the fatality rate in a timely manner.

The proposed test, together with the real-time fatality rate estimator, can help to assess the effectiveness of existing health policies and interventions and to decide the need of new treatments especially in the absence of large scale clinical trials for a newly emerging epidemic. The setup of the problem, the proposed test, and its asymptotic properties will be discussed in the next section. The proposed method is applied to Hong Kong and Beijing SARS data in Section 3. Results of a simulation study to assess the performance of the test are reported in Section 4.

2. A Test for Constant Fatality Rate

Consider a population initially consisting of n healthy individuals subject to infection during an epidemic. When an individual is infected, he/she will be admitted to and treated in a hospital some time after the incubation period, usually longer at the beginning of the epidemic and shorter after the epidemic has taken off when the community is more aware of the disease. Inpatients are assumed to either die at a rate of $\gamma_1(t)$ or recover at a rate of $\gamma_2(t)$ on day t, where t is the calendar time since the beginning of the epidemic rather than the time since individual onset as in traditional survival analysis. By focusing on the calendar time instead of the survival time, we gain more insight in the trend of the fatality rate at the population level over the course of the epidemic. The population real time or time-varying fatality rate is defined as

$$\pi(t) = \frac{\gamma_1(t)}{\gamma_1(t) + \gamma_2(t)}, \tag{1}$$

which can be viewed as the conditional probability of death given that an event of death or recovery occurs at time t. This fatality rate is a measure of the combined effect of the processes of death and recovery which is similar to the setting of the competing risks model (Yip and Lam, 1992; Kalbfleisch and Prentice, 2002, Chapter 7).

The data consist of the processes $\{N_1(t), N_2(t)\}$, the cumulative numbers of deaths and recoveries up to time t and $\{I(t)\}$, the number of inpatients just before time t, $0 \le t \le \tau$, where τ is a prespecified time measured in days. We decide to observe the process till time τ in the hope that τ is large enough to cover the entire anticipated length of the epidemic. Now suppose that $N_1(t)$ and $N_2(t)$ are counting processes with intensity processes $\gamma_1(t)$ and $\gamma_2(t)$, respectively, that satisfy the multiplicative intensity model

$$\Pr\{dN_i(t) = 1 \,|\, \mathcal{F}_{t-}\} = \gamma_i(t)I(t)\,dt \quad i = 1, 2,$$

where $\mathcal{F}_t = \{I(u), N_i(u), i = 1, 2; u \in [0, t]\}$ is the σ-field representing the history of the epidemic process up to time t that satisfies the usual regularity conditions (Andersen et al., 1993, p. 60). Define

$$M_i(t) = N_i(t) - \int_0^t \gamma_i(s)I(s)\,ds, \quad i = 1, 2.$$

These are zero mean martingales with respect to the filtration $\{\mathcal{F}_{t-}, t \ge 0\}$ (Andersen et al., 1993, pp. 72–82). Also, define $J(s) = 1\{I(s) > 0\}$, the indicator function of having at least one inpatient at time $s-$. We assume that no two events occur simultaneously. It is of interest to test the null hypothesis that the fatality rate $\pi(t)$ as defined in (1) is constant over the period $[0, \tau]$ against the alternative hypothesis that $\pi(t)$ decreases with t. We note that $\gamma_1(t)/\gamma_2(t) = 1/\{1/\pi(t) - 1\}$. Thus we seek a statistical test for testing the equivalent hypothesis

$$H_0 : \frac{\gamma_1(t)}{\gamma_2(t)} = c_0, \quad \text{against} \quad H_1 : \frac{\gamma_1(t)}{\gamma_2(t)} \downarrow \text{ in } t, \quad 0 \le t \le \tau,$$

where c_0 is an unknown constant. Let

$$Z_n(t) = \frac{1}{\sqrt{n}} \left\{ \int_0^t \frac{J(s)N_1(s-)}{I(s)}\,dN_2(s) - \int_0^t \frac{J(s)N_2(s-)}{I(s)}\,dN_1(s) \right\}. \tag{2}$$

Note that there is no contribution to $Z_n(t)$ for the period during which $J(s) = 0$. Define $\Lambda_i(t) = \int_0^t \gamma_i(s)I(s)\,ds$ and $\Gamma_i(t) = \int_0^t \gamma_i(s)\,ds$. Using integration by parts, $Z_n(t)$ may be rewritten as

$$Z_n(t) = \frac{1}{\sqrt{n}} \left\{ \int_0^t \frac{J(s)N_1(s-)}{I(s)}\,dM_2(s) \right.$$
$$\left. + \int_0^t \frac{M_1(s-) + \Lambda_1(s)}{I(s)} J(s)\gamma_2(s)I(s)\,ds \right\}$$
$$- \frac{1}{\sqrt{n}} \left\{ \int_0^t \frac{J(s)N_2(s-)}{I(s)}\,dM_1(s) \right.$$
$$\left. + \int_0^t \frac{M_2(s-) + \Lambda_2(s)}{I(s)} J(s)\gamma_1(s)I(s)\,ds \right\}$$
$$= \frac{1}{\sqrt{n}} \left[\int_0^t \left\{ \frac{J(s)N_1(s-)}{I(s)} - (\Gamma_1(t) - \Gamma_1(s)) \right\} dM_2(s) \right.$$
$$\left. - \int_0^t \left\{ \frac{J(s)N_2(s-)}{I(s)} - (\Gamma_2(t) - \Gamma_2(s)) \right\} dM_1(s) \right]$$
$$+ \frac{1}{\sqrt{n}} \left\{ \int_0^t \Lambda_1(s)\gamma_2(s)\,ds - \int_0^t \Lambda_2(s)\gamma_1(s)\,ds \right\}. \tag{3}$$

The last term in (3) is 0 under H_0 and $Z_n(t)$ is strictly positive under H_1. Hence, the proposed test statistic is consistent against any decreasing fatality rate. We assume that

(i) $\lim_{n\to\infty} \frac{N_1(t-)}{n} = p_1(t)$ in probability,

(ii) $\lim_{n\to\infty} \frac{N_2(t-)}{n} = p_2(t)$ in probability,

(iii) $\lim_{n\to\infty} \frac{I(t)}{n} = p_I(t) > 0$ in probability,

for $0 < t < \tau$. Using the Rebelledo martingale central limit theorem (Andersen et al., 1993, p. 83), we get that as $n \to \infty$ and under H_0, the process $Z_n = \{Z_n(t), 0 \le t \le \tau\}$ converges weakly to a process $U = \{U(t), 0 \le t \le \tau\}$ in $D[0, \tau]$ with the Skorohod topology, where $\{U(t), 0 \le t \le \tau\}$ is a continuous zero-mean Gaussian martingale and $\text{cov}\{U(s), U(t)\} = \sigma^2\{\min(t, s)\}$ with

$$\sigma^2(t) = \int_0^t \left\{ \frac{p_1(s)}{p_I(s)} - (\Gamma_1(t) - \Gamma_1(s)) \right\}^2 p_I(s)\gamma_2(s)\, ds$$

$$+ \int_0^t \left\{ \frac{p_2(s)}{p_I(s)} - (\Gamma_2(t) - \Gamma_2(s)) \right\}^2 p_I(s)\gamma_1(s)\, ds.$$

Let

$$\hat{\sigma}_n^2(t) = \frac{1}{n} \int_0^t \left\{ \frac{J(s)N_1(s-)}{I(s)} - (\hat{\Gamma}_1(t) - \hat{\Gamma}_1(s)) \right\}^2 dN_2(s)$$

$$+ \frac{1}{n} \int_0^t \left\{ \frac{J(s)N_2(s-)}{I(s)} - (\hat{\Gamma}_2(t) - \hat{\Gamma}_2(s)) \right\}^2 dN_1(s),$$

where $\hat{\Gamma}_i(t) = \int_0^t \frac{J(s)}{I(s)} dN_i(s)$ is just the usual Nelson–Aalen estimator for $\Gamma_i(t)$. Then, one can easily show that

$$\sup_{t\in[0,\tau]} \left| \hat{\sigma}_n^2(t) - \sigma^2(t) \right| \to 0 \text{ in probability} \qquad (4)$$

as $n \to \infty$ using the same theorem.

At some prespecified time τ, we can establish a straightforward test statistic given by

$$V_n(\tau) = \frac{Z_n(\tau)}{\hat{\sigma}_n(\tau)}. \qquad (5)$$

Under the usual mild regularity conditions, the statistic $V_n(\tau)$ has a standard normal distribution asymptotically under the null hypothesis. The performance of the test is very good in terms of unbiasedness and the power of the test. However, the test statistic in (5) does not possess the desirable property of being able to reject the null hypothesis as soon as we have accumulated enough statistical evidence, but the decision can only be made at time τ.

Here, $\hat{\sigma}_n^2(t)$ is not necessarily a nondecreasing function even though a decrement in $\hat{\sigma}_n^2(t)$ is very unlikely empirically. To get around this problem, we simply let $h_n(s) = \inf\{t \ge 0 \,|\, \hat{\sigma}_n^2(t) \ge s\}$ and $h(s) = \inf\{t \ge 0 \,|\, \sigma^2(t) \ge s\}$. Note that h is a nondecreasing, continuous, nonrandom function on $[0, \sigma^2(\tau)]$. It follows from (4) that $h_n(s)$ converges in probability to $h(s)$ uniformly for $s \in [0, \sigma^2(\tau)]$ as $n \to \infty$ and that (Z_n, h_n) converges weakly to (U, h) in $D[0, \tau] \times D[0, \sigma^2(\tau)]$ (Billingsley, 1999, Theorem 3.9). Using an argument similar to Billingsley (1999, p. 151), it can be shown that the process $Z_n(h_n) = \{Z_n(h_n(s)), 0 \le s \le \hat{\sigma}_n^2(\tau)\}$ converges weakly to the Gaussian process $\{U(h(s)), 0 \le s \le \sigma^2(\tau)\}$ as $n \to \infty$. Further,

$$\text{cov}\{U(h(s_1)), U(h(s_2))\} = \sigma^2(\min(h(s_1), h(s_2))) = \min(s_1, s_2),$$

implying that the process $\{U(h(s)), 0 \le s \le \sigma^2(\tau)\}$ is a Brownian motion.

Technically, for the above results to hold, we require that no two events occur simultaneously and the processes $N_1(t)$, $N_2(t)$, and $I(t)$ are observed continuously. However, in practice, the observations are usually taken on a daily basis and the exact transition times are generally unobservable. In such situations, we propose to randomly impute the death and discharge (recovery) times within the day, or simply impute the order of deaths and discharges on that day because the test statistic only requires the order of occurrences of the events rather than the exact time to occurrences of the events. Also, we assume that all inpatients enter the system at the beginning of each day. From our empirical experience, the effect of randomization of the sequence of the deaths and discharges within a day has minimal effect on the behavior of the proposed test statistic.

As the processes were observed on a daily basis in practice, for ease of presentation, we define a partition $0 = s_0 \le s_1 \le s_2 \le \cdots \le s_\tau$ where $s_t = \hat{\sigma}_n^2(t)$ if $\hat{\sigma}_n^2(t) > \hat{\sigma}_n^2(t-1)$ or else $s_t = \hat{\sigma}_n^2(t-1)$ for $t = 1, \ldots, \tau$. A sequential test for constant fatality rate can then be based on $\{Z_n(h_n(s_1)), \ldots, Z_n(h_n(s_\tau))\}$. The idea is to compute a statistic $Z_n(t)$ at the end of the tth day and compare it with its corresponding stopping boundary c_t for $t = 1, \ldots, \tau$. The null hypothesis of constant fatality rate can be rejected at the end of day t if $Z_n(t)$ exceeds c_t for the first time.

The set of appropriate stopping boundaries $(c_1, \ldots, c_\tau)$ are chosen such that

$$\alpha = 1 - P\big[\cap_{t=1}^\tau \{Z_n(h_n(s_t)) < c_t\} \big],$$

in order to preserve the overall probability of committing a type I error at the prespecified level α (Armitage, McPherson, and Rowe, 1969). Adjustments are made on the level of significance for each comparison in the repeated significance tests using the α-spending function idea as proposed in Lan and DeMets (1983). A monotone increasing function of t for $t = 0, 1, \ldots, \tau$ with $\alpha(0) = 0$ and $\alpha(\tau) = \alpha$ is specified. Let $T = \min\{t : Z_n(h_n(s_t)) > c_t\}$. A popular choice of the α-spending function is

$$\alpha(t) = 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t/\tau}} \right),$$

proposed by Lans and DeMets (1983), where $\Phi(.)$ is the distribution function of a standard normal random variable and z_α satisfies $\Phi(z_\alpha) = 1 - \alpha$. This α-spending function will be used throughout this article.

The probability that H_0 is rejected on the tth day is then

$$P(T = t) = P(Z_n(h_n(s_t)) > c_t; Z_n(h_n(s_i))$$

$$< c_i, \ 1 \le i \le t - 1).$$

With the set of $\alpha(t)$'s obtained from the α-spending function, we have the recursive relationship

$$\alpha(1) = P(T = 1) = P(Z_n(h_n(s_1)) > c_1),$$

$$\alpha(2) = \alpha(1) + P(T = 2)$$

$$= \alpha(1) + P\{Z_n(h_n(s_2)) > c_2; Z_n(h_n(s_1)) < c_1\},$$

$$\alpha(j) = \alpha(j - 1) + P(T = j) \quad \text{for} \quad j = 3, \ldots, \tau.$$

As the process $Z_n(h_n(s))$ converges in distribution to a Brownian motion process, the asymptotic joint distribution of $\{Z_n(h_n(s_1)), \ldots, Z_n(h_n(s_\tau))\}$ is Gaussian with $\text{Cov}\{Z_n(h_n(s_i)), Z_n(h_n(s_j))\} = \min(s_i, s_j)$ for all $s_1, \ldots, s_\tau$, and that the increments $Z_n(h_n(s_j)) - Z_n(h_n(s_{j-1}))$ ($j = 1, \ldots, \tau$) are independent. The values of the stopping boundaries $c_1, \ldots, c_k$ can thus be obtained recursively through

$$P(T = k) = \alpha(k) - \alpha(k-1)$$

$$= \int_{c_k}^{\infty} \int_{-\infty}^{c_{k-1}} \cdots \int_{-\infty}^{c_1} \prod_{t=1}^{k} \frac{1}{\sqrt{2\pi(s_t - s_{t-1})}}$$

$$\times \exp\left\{-\frac{(u_t - u_{t-1})^2}{2\pi(s_t - s_{t-1})}\right\} du_1 \, du_2 \cdots du_k$$

numerically, with $u_0 = 0$.

The value of the stopping boundary c_t is largely affected by the value of s_t. Hence the standardized statistic $V_n(t) = Z_n(h_n(s_t))/(s_t)^{\frac{1}{2}}$ is considered in the following discussion and will be compared with the new stopping boundaries given by $b_t = c_t/(s_t)^{\frac{1}{2}}$ so that a plot of the $V_n(t)$ and b_t over t will be smoother. In the cases with $s_t = s_{t-1}$ for $t = 1, \ldots, \tau$, the value of $Z_n(h_n(s_t))$ will remain unchanged, and we will simply skip through the comparison at the end of day t, move forward to day $t+1$ and compare $V_n(t+1)$ with b_{t+1} where b_{t+1} satisfies $P(T = t+1) = \alpha(t+1) - \alpha(t-1)$ provided that $s_{t+1} > s_{t-1}$.

3. Applications

The proposed test is applied to Hong Kong and Beijing SARS data. There were 1755 SARS cases in total, with 299 deaths in Hong Kong (Department of Health, HKSAR, 2003) and 2521 cases with 193 deaths in Beijing (Beijing Center for Disease Control and Prevention, 2003). The daily numbers of deaths, recoveries and inpatients in the two regions are used. For Hong Kong, the data available spanned from March 12, 2003 to June 30, 2003 with a total of 110 days. For Beijing, the data available spanned from April 21, 2003 to July 1, 2003, with a total of 72 days. There were only 29 and 27 inpatients left in the hospitals, respectively, for Hong Kong and Beijing at the end of the periods as specified above. When there are more than one death or discharge on the same day, we impute the rank of the death and discharge times within the same day at random. For simplicity, we set $\tau = 60$ for both cases and the overall level of significance is set at $\alpha = 0.05$.

For easy reference, we plot the standardized statistics $V_n(t)$ (represented by the star) and the stopping boundaries b_t (represented by the solid line) against the corresponding day for the Hong Kong and Beijing data in the left and right panels of Figure 1, respectively. The null hypothesis of constant fatality is not rejected for Hong Kong data throughout the said period. The same conclusion was drawn based on the proposed sequential tests with $\tau = 110$ and the test statistic $V_n(\tau)$ in (5). However, the null hypothesis is rejected on day 22 (May 12, 2003) for Beijing data. In other words, the SARS fatality rate in Hong Kong was concluded to be constant throughout the period, while that in Beijing showed a decreasing trend in fatality over time, which could possibly be due to their improvement of treatment including the use

of Chinese medicine. In Beijing, they had also implemented very strict quarantine measures in controlling the infections of SARS and a specialized hospital was built for SARS patients in seven days and was opened on May 1, 2003 (Pang et al., 2003) that helped to control the spread of the epidemic in the hospital settings because most of the hospitals are not well equipped to deal with the SARS patients and infection within hospital would become serious.

In the case of constant fatality, a simple estimator for $\pi = \pi(t)$ at time t is given by

$$\hat{\pi}(t) = \frac{N_1(t)}{N_1(t) + N_2(t)} \quad \text{for } 0 < t \le \tau,$$

which is shown to be the MLE in the Appendix. The estimated $\hat{\pi}(t)$ for the constant fatality rate and the estimated time-varying fatality rate $\tilde{\pi}(t)$ of Yip, Lam, et al. (2005) over t were obtained for the Hong Kong and Beijing data. The estimates and their associated 95% pointwise confidence limits were plotted in the left and right panels of Figure 2, respectively. The above conclusion that the null hypothesis of constant fatality rate is not rejected for the Hong Kong data is further supported by the plot in the left panel of Figure 2. The estimate based on $\hat{\pi}(t)$ has been very stable since April 22, while the estimated time-varying fatality rate $\tilde{\pi}(t)$ simply fluctuates around $\hat{\pi}(t)$ over t. The two sets of estimates agree closely with each other in general. The 95% pointwise confidence intervals based on $\hat{\pi}(t)$ are much narrower than those based on the time-varying fatality estimates, $\tilde{\pi}(t)$. Hence the simple estimator $\hat{\pi}(t)$ should be more efficient if the constant fatality assumption is valid. The plot in the right panel of Figure 2 shows that the constant fatality is highly unlikely to be the case for Beijing data. The patterns of the two estimates differ considerably. The estimate based on $\hat{\pi}(t)$ does not appear to be constant itself, but exhibits a decreasing trend over time. However, it has a similar pattern as $\tilde{\pi}(t)$ but the value of $\hat{\pi}(t)$ is consistently about 10–13% greater than that of $\tilde{\pi}(t)$. It is mainly due to the inertia of the MLE which is strongly influenced by the higher number of deaths in the early part of the epidemic. In Beijing, there were more confirmed cases after May 1, but with fewer deaths afterward. The proposed test statistic for constant fatality seems to work well in these applications. A large scale simulation study is reported to investigate the performance of the proposed test empirically in the next section.

4. Simulation Study

We simulate several scenarios to assess the performance of the proposed test. In the simulation we assume that data with exact death and recovery times are not available. This mimics the practical situation that only the daily counts are available. Because only the death and recovery processes but not the infection process determine the test statistic, simulating the infection times is unnecessary while our focus is on the fatality rate. Therefore, we simply make use of the observed daily number of inpatients as in Hong Kong SARS epidemic in the simulation. The daily numbers of deaths and recoveries are simulated according to a multinomial distribution with probabilities $q_1(t)$ and $q_2(t)$ on day t, respectively. The value of τ is assumed to be 60 days, and the overall level of significance is set at $\alpha = 0.05$ throughout. In each set of simulations,
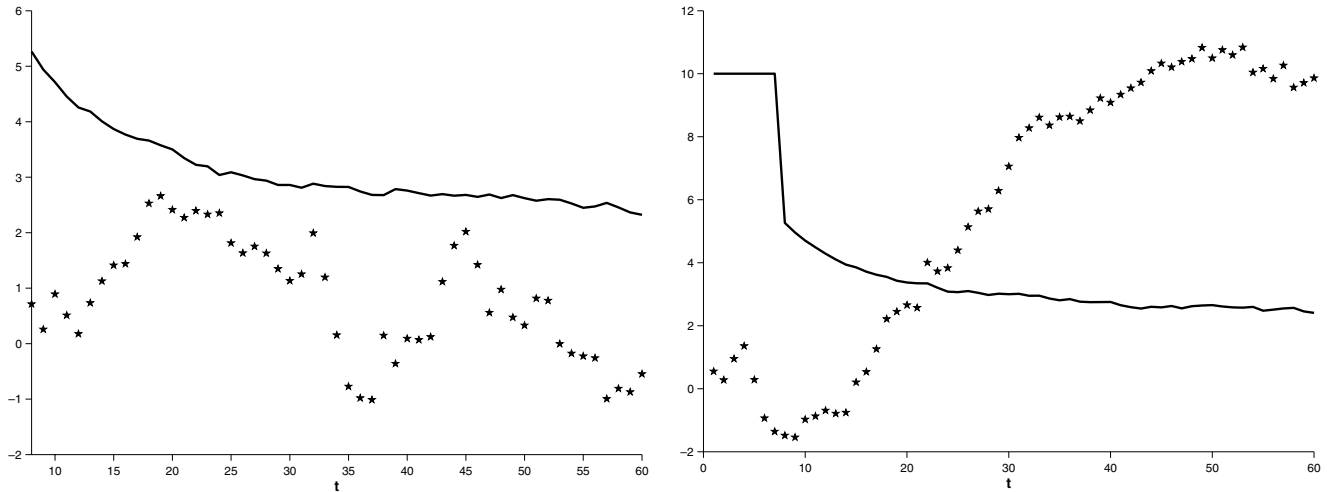
Figure 1. Plot of the standardized statistic $V_n(t)$ (star) and the stopping boundaries (solid line) against the corresponding day for outbreak of SARS in Hong Kong (left panel) and Beijing (right panel).
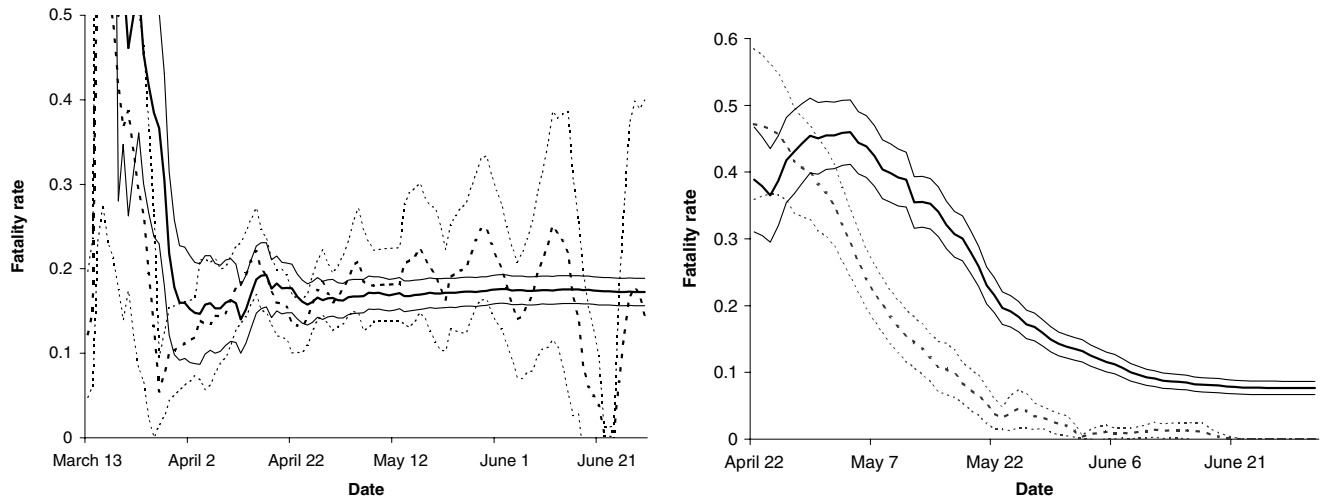

Figure 2. The estimated time-varying fatality rate (dashed line) and $\hat{\pi}(t)$ for the constant fatality rate (solid line) and their corresponding 95% pointwise confidence limits for outbreak of SARS in Hong Kong (left panel) and Beijing (right panel).

10,000 data sets are generated and the proportion of data sets with the null hypothesis being rejected is the empirical size or power. Results are tabulated in Tables 1 and 2.

Different scenarios with constant fatality rate, that is the null hypothesis, namely (A) constant death and recovery probabilities; (B) linearly decreasing/increasing death probability with proportional recovery probability; and (C) exponentially decreasing death probability with proportional recovery probability, are simulated to assess the empirical significance level of the test. Table 1 shows that the empirical sizes of the test are very close to $\alpha = 0.05$ in all cases. The proposed test is seen to be nearly unbiased.

Different scenarios with decreasing fatality rate, namely (A) constant death probability and piecewise constant recovery probability with a sudden increment at day t_0; (B) constant death probability and linearly increasing recovery probability; (C) linearly decreasing death probability and

constant recovery probability; and (D) linearly increasing recovery probability and linearly increasing/decreasing death probability, are simulated to assess the empirical power of the proposed test. The simulation results are tabulated in Table 2. It is obvious that the empirical powers are very high (over 90%) in general unless the decrement in the fatality rate is extremely slow like the last case of scenario (C). From the results based on scenario (A), we may conclude that the test is extremely sensitive to the changes in the fatality rate and the null hypothesis may be rejected as early as within a week since the sudden drop in the fatality rate on the average. The column $1 - \pi(\tau)/\pi(0)$ represents the relative decrement in the fatality rate from the start to time τ or the end of the epidemic. The power of the test increases with the relative decrement. Moreover the empirical powers remain to be quite high in general even when the relative decrement is only around 20%. Considering all scenarios, the proposed test

Table 1
Simulation results for the empirical significance levels of the proposed test under different scenarios: (A)
constant death and recovery probabilities; (B) *linearly decreasing/increasing death probability with proportional
recovery probability; and* (C) *exponentially decreasing probability with proportional recovery probability.*
*avg.$N_1(\tau)$ and avg.$N_2(\tau)$ indicate the averages over all the simulations of the total numbers of deaths and
recoveries, respectively. avg.rej.day represents the mean number of days (from the origin) on which the null
hypothesis is first rejected among all the rejected cases.*

Scenario	$q_1(t)$	$q_2(t)$	avg.$N_1(\tau)$	avg.$N_2(\tau)$	π	Empirical size (%)
(A)	0.008	0.01	273.3646	341.7217	0.4444	4.93
(A)	0.008	0.02	273.2267	683.2609	0.2857	4.91
(A)	0.008	0.04	273.3955	1367.3130	0.1667	4.99
(A)	0.008	0.08	273.4382	2735.7120	0.0909	5.13
(A)	0.01	0.01	341.6004	341.6844	0.5000	4.78
(A)	0.01	0.02	341.4107	683.5364	0.3333	4.90
(A)	0.01	0.04	341.9348	1367.5020	0.2000	4.97
(A)	0.005	0.01	170.7891	341.5622	0.3333	4.94
(A)	0.005	0.05	170.9705	1709.1450	0.0909	5.20
(B)	$0.005 + 0.0001t$	$q_1(t)$	294.9111	294.6624	0.5000	4.69
(B)	$0.005 + 0.0001t$	$2\ q_1(t)$	294.9808	589.2811	0.3333	4.87
(B)	$0.005 + 0.0001t$	$4\ q_1(t)$	295.0944	1179.5370	0.2000	5.06
(B)	$0.005 + 0.0001t$	$8\ q_1(t)$	295.0426	2360.1870	0.1111	5.18
(B)	$0.1 - 0.0008t$	$q_1(t)$	2426.8190	2426.9260	0.5000	5.09
(B)	$0.1 - 0.0008t$	$2\ q_1(t)$	2426.7940	4853.7840	0.3333	4.83
(B)	$0.1 - 0.0008t$	$4\ q_1(t)$	2426.4980	9706.9090	0.2000	5.07
(B)	$0.1 - 0.0008t$	$8\ q_1(t)$	2426.6430	19415.1500	0.1111	5.04
(C)	$0.01 \exp(-0.01t)$	$2\ q_1(t)$	239.4667	479.4933	0.3333	4.96
(C)	$0.01 \exp(-0.01t)$	$8\ q_1(t)$	239.7724	1918.8120	0.1111	5.36
(C)	$0.02 \exp(-0.01t)$	$2\ q_1(t)$	479.3387	958.8171	0.3333	4.90
(C)	$0.02 \exp(-0.01t)$	$8\ q_1(t)$	479.6293	3836.9340	0.1111	5.17
(C)	$0.1 \exp(-0.04t)$	$4\ q_1(t)$	916.0915	3664.2280	0.2000	5.11
(C)	$0.2 \exp(-0.02t)$	$4\ q_1(t)$	3422.5490	13689.1600	0.2000	5.05

statistic seems to be very powerful in general, and is sensitive enough to pick up various forms of changes in the fatality rate. Therefore, the performance of the proposed test is very satisfactory and will be useful to monitor the fatality rate of a newly emerging disease.

5. Discussion

We have presented a testing procedure which is able to detect a decreasing fatality rate in a macro perspective, especially when the decrease in death probability or increase in recovery probability happens within a more intense period in the midst of an epidemic. When an effective new treatment or an intervention policy is implemented, it should cause a marked drop in the fatality rate. By formulating a model with time-varying fatality rate, the proposed test is a useful tool to justify or to evaluate the effectiveness of the adopted health policies and/or the clinical management. A main advantage of this test is that as information accumulates over time, we are able to reject the null hypothesis as soon as there is enough evidence of decreasing fatality over time. Hence, more effective intervention measures can be launched in place to reduce fatality.

Applications of $V_n(\tau)$ to the Hong Kong (with $\tau = 110$) and Beijing data (with $\tau = 72$) lead to essentially the same conclusions with asymptotic p-values 0.3082 and < 0.0001, respectively. The performance of the test statistic $V_n(\tau)$ is also studied in a simulation study not reported here. The performance is very similar to the one proposed, but the shortcoming of the test $V_n(\tau)$ is that it is evaluated at a prespecified time τ, but we are not able to reject H_0 before time τ even though there exhibits strong statistical evidence. However, this could also be the merit in some other applications. Modification of the proposed sequential tests can be made easily by using different times of interim analysis, say by making a comparison every three days instead of every day. This version has the advantage that the environmental and weather conditions will be more homogeneous within each period under consideration so that some uncontrollable variation can be grouped together or marginalized as the fatality rate may also be affected by the temperature or humidity. More work needs to be done to explore the feasibility of the modified test in practice.

As of January 29, 2007 (WHO, 2007), there are a total of 164 deaths out of 270 human cases of H5N1 avian influenza, giving an estimated traditional fatality rate greater than 50%. If the strain acquires human-to-human transmissibility, its virulence to humans may or may not be the same as the historical level when transmission is limited to avian-to-human only. In the case of an outbreak due to human-to-human transmission, the proposed test will then be an extremely useful tool to provide a formal statistical method to test the hypothesis of decreasing fatality presumably due to some effective

Table 2
Simulation results for the empirical powers of the proposed test under different scenarios: (A) stepwise increase in recovery probability; (B) constant death probability and linearly increasing recovery probability; (C) linearly decreasing death probability and constant recovery probability; and (D) linearly increasing recovery probability and linearly increasing/decreasing death probability. avg.$N_1(\tau)$ and avg.$N_2(\tau)$ indicate the averages over all the simulations of the total numbers of deaths and recoveries, respectively. avg.rej.day represents the mean number of days (from the origin) on which the null hypothesis is first rejected among all the rejected cases.

Scenario	$q_1(t)$	$q_2(t)$	avg.$N_1(\tau)$	avg.$N_2(\tau)$	$\pi(0)$	$1 - \frac{\pi(\tau)}{\pi(0)}$	Power (%)	avg.rej.day
(A)	0.008	$0.04 + 0.04\ I(t \geq 30)$	274.0853	2315.7870	0.1667	0.4545	98.55	35.7
(A)	0.008	$0.04 + 0.04\ I(t \geq 40)$	274.1753	1944.7080	0.1667	0.4545	99.64	44.0
(A)	0.008	$0.04 + 0.04\ I(t \geq 50)$	273.3781	1608.6220	0.1667	0.4545	95.39	53.5
(A)	0.008	$0.04 + 0.04\ I(t \geq 55)$	273.6446	1484.8250	0.1667	0.4545	61.16	57.3
(A)	0.01	$0.02 + 0.02\ I(t \geq 30)$	341.7206	1157.8750	0.3333	0.4000	98.29	35.8
(A)	0.01	$0.02 + 0.02\ I(t \geq 40)$	342.1126	971.2872	0.3333	0.4000	99.83	44.0
(A)	0.01	$0.02 + 0.02\ I(t \geq 50)$	341.5698	803.1143	0.3333	0.4000	94.85	54.1
(A)	0.01	$0.02 + 0.02\ I(t \geq 55)$	341.8031	742.0484	0.3333	0.4000	58.74	57.3
(A)	0.01	$0.05 + 0.04\ I(t \geq 30)$	341.9205	2657.4600	0.1667	0.4000	97.25	36.0
(A)	0.01	$0.05 + 0.04\ I(t \geq 40)$	341.7803	2286.5690	0.1667	0.4000	99.48	44.6
(A)	0.01	$0.05 + 0.04\ I(t \geq 50)$	341.6964	1950.2510	0.1667	0.4000	92.24	54.3
(A)	0.01	$0.05 + 0.04\ I(t \geq 55)$	341.9452	1827.1990	0.1667	0.4000	52.05	57.2
(B)	0.05	$0.05 + 0.0008t$	1709.7360	2703.1530	0.5000	0.3243	98.52	40.3
(B)	0.05	$0.05 + 0.001t$	1710.1950	2951.0670	0.5000	0.3750	99.89	37.3
(B)	0.05	$0.05 + 0.004t$	1710.2290	6677.1620	0.5000	0.7059	100.00	26.0
(B)	0.005	$0.005 + 0.002t$	171.1493	2653.7260	0.5000	0.9230	99.06	37.6
(B)	0.005	$0.005 + 0.004t$	171.2511	5136.4080	0.5000	0.9600	99.51	36.6
(B)	0.005	$0.005 + 0.008t$	171.2199	10104.5500	0.5000	0.9796	99.48	35.9
(B)	0.005	$0.005 + 0.01t$	171.2208	12587.5500	0.5000	0.9836	99.42	35.8
(C)	$0.05 - 0.0004t$	0.05	1212.7220	1709.3670	0.5000	0.3158	96.92	44.3
(C)	$0.05 - 0.0002t$	0.05	1460.6430	1708.7680	0.5000	0.1364	42.13	48.5
(C)	$0.05 - 0.0001t$	0.05	1585.4610	1709.6150	0.5000	0.0638	15.36	48.4
(D)	$0.005 + 0.0001t$	$0.01 + 0.005t$	294.9207	6550.5110	0.3333	0.9314	91.40	40.5
(D)	$0.02 + 0.001t$	$0.02 + 0.008t$	1925.6180	10617.9800	0.5000	0.7241	99.11	33.8
(D)	$0.1 - 0.001t$	$0.05 + 0.005t$	2178.4020	7919.4180	0.6667	0.7805	100.00	20.5
(D)	$0.3 - 0.004t$	$0.2 + 0.005t$	5295.1120	13048.2300	0.6000	0.8214	100.00	20.3

interventions in a sense that the alternative hypothesis can be accepted as soon as there is enough statistical evidence so that prompt and decisive action can be taken more effectively to suppress the fatality at the earliest time. A delay in the implementation of some effective intervention may result in more life loss globally. Nevertheless, we hope this test will never be used in practice.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Armitage, P., McPherson, C. K., and Rowe, B. C. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society, Series A* **132,** 235–244.

Beijing Center for Disease Control and Prevention. (2003). Latest figures on severe acute respiratory syndrome in Mainland China. August 16, 2003. (Available from http://www.bjcdc.org/jkxw/shownews1.asp?newsid=996)

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edition. New York: Wiley.

Department of Health of Hong Kong Government. (2003). Latest figures on severe acute respiratory syndrome. (Available from http://www.info.gov.hk/dh/diseases/ap/eng/infected.htm)

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. Hoboken, New Jersey: Wiley.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70,** 659–663.

Last, J. M. (1995). *A Dictionary of Epidemiology*, 3rd edition. New York: Oxford University Press.

Pang, X. H., Zhu, Z. H., Xu, F. J., Guo, J. Y., Gong, X. H., Liu, D. L., Liu, Z. J., Chin, D. P., and Feikin, D. R. (2003). Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing. *JAMA* **290,** 3215–3221.