PAPER ID: PMC7109660

TITLE:
A Chain Multinomial Model for Estimating the Real-Time Fatality Rate of a Disease, with an Application to Severe Acute Respiratory Syndrome

ABSTRACT:
It is well known that statistics using cumulative data are insensitive to changes. World Health Organization (WHO) estimates of fatality rates are of the above type, which may not be able to reflect the latest changes in fatality due to treatment or government policy in a timely fashion. Here, the authors propose an estimate of a real-time fatality rate based on a chain multinomial model with a kernel function. It is more accurate than the WHO estimate in describing fatality, especially earlier in the course of an epidemic. The estimator provides useful information for public health policy makers for understanding the severity of the disease or evaluating the effects of treatments or policies within a shorter time period, which is critical in disease control during an outbreak. Simulation results showed that the performance of the proposed estimator is superior to that of the WHO estimator in terms of its sensitivity to changes and its timeliness in reflecting the severity of the disease.

THE ESTIMATOR:
We take the time periods to be days. Time is measured in the number of days since the first case of the disease was diagnosed or reported. Suppose that we observe the numbers N_{1t} and N_{2t} of the H_{t-1} inpatients at the start of day t who respectively die from the disease or recover on day t. Note that the number of inpatients H_{t-1} includes new SARS admissions on day $t-1$ as well as existing inpatients who neither die nor recover. On day t, an inpatient dies or recovers with probabilities p_{1t} and p_{2t}, respectively. Using the information on the number of inpatients at the end of day $t-1$, $N_{1t}, N_{2t}|H_{t-1} \sim$ multinomial $(H_{t-1}, p_{1t}, p_{2t})$ (see Becker (7) and Yip and Huggins (8)). Consequently, N_{it} and N_{js} are conditionally uncorrelated for $s \neq t$, $i, j = 1, 2$. Simple estimators for p_{1t} and p_{2t} are N_{1t}/H_{t-1} and N_{2t}/H_{t-1}, respectively. The estimated variances of $\hat{p}_{it}$ are given by $\hat{p}_{it}(1-\hat{p}_{it})/H_{t-1}$ for $i = 1, 2$, and the covariance between $\hat{p}_{1t}$ and $\hat{p}_{2t}$ is $-\hat{p}_{1t}\hat{p}_{2t}/H_{t-1}$. A simple estimator for the fatality rate at time t, π_t, is

$$\mathrm{\hat{\pi}}_{t} = \frac{\hat{p}_{1t}}{\hat{p}_{1t}+\hat{p}_{2t}}.$$

A plot of these unsmoothed estimates for the Hong Kong data is given in figure 1. These daily estimates are highly variable, and there is no long-term trend in the estimated fatality rates. The daily estimates are generally larger than the WHO estimates, which increase throughout the course of the epidemic. We propose to reduce the variability and hence clarify the long-term trend by computing smooth estimates.

For computation of smooth estimates of the recovery and death rates for SARS, a kernel method was applied to $\hat{p}_{1t}$ and $\hat{p}_{2t}$. Let $K_b(x) = b^{-1}K(x/b)$ be a kernel function of bandwidth b, with support in $[-1.1]$. The Nadaraya-Watson estimator (9, 10) for p_{it} is given by

$$\tilde{p}_{it} = \frac{\sum_{s=1}^{T}K_{b}(t-s)\hat{p}_{is}}{\sum_{s=1}^{T}K_{b}(t-s)},\,i=1,\,2,$$

where T is the last day of the epidemic. Because $\hat{p}_{it}$ and $\hat{p}_{is}$ are conditionally uncorrelated for $s \neq t$, the variances of the $\tilde{p}_{it}$, $i = 1, 2$, are given by

$$\mathrm{var}(\tilde{p}_{it}) = \frac{\sum_{s=1}^{T}K_{b}^{2}(t-s)\hat{p}_{is}(1-\hat{p}_{is})/H_{s-1}}{\left(\sum_{s=1}^{T}K_{b}(t-s)\right)^{2}}$$

and the covariance of $\tilde{p}_{1t}$ and $\tilde{p}_{2t}$ is

$$\mathrm{cov}(\tilde{p}_{1t},\, \tilde{p}_{2t}) = \frac{-\sum_{s=1}^{T}K_{b}^{2}(t-s)\hat{p}_{1s}\hat{p}_{2s}/H_{s-1}}{\left(\sum_{s=1}^{T}K_{b}(t-s)\right)^{2}}.$$

An estimator for the ratio of the death and recovery transition probabilities is given by $\tilde{\theta}_t = \tilde{p}_{1t}/\tilde{p}_{2t}$, and its variance can be obtained by the standard δ method (11, p. 388):

$$\mathrm{var}$$

$$\mathrm{var}(\tilde{\theta}_t) = \frac{1}{\tilde{p}_{2t}^2}\mathrm{var}(\tilde{p}_{1t}) + \frac{\tilde{p}_{1t}^2}{\tilde{p}_{2t}^4}\mathrm{var}(\tilde{p}_{2t}) - \frac{2\tilde{p}_{1t}}{\tilde{p}_{2t}^3}\mathrm{cov}(\tilde{p}_{1t},\tilde{p}_{2t}).$$

Then, we estimate π_t by

$$\tilde{\pi}_t = \frac{\tilde{p}_{1t}}{\tilde{p}_{1t} + \tilde{p}_{2t}} = \frac{\tilde{\theta}_t}{1 + \tilde{\theta}_t}.$$

The δ method yields a variance estimator for $\tilde{\pi}_t$ as

$$\mathrm{var}(\tilde{\pi}_t) = \frac{1}{(1 + \tilde{\theta}_t)^4}\mathrm{var}(\tilde{\theta}_t)$$

and approximate $(1 - \alpha) \times 100$ percent pointwise confidence limits for $\tilde{\pi}_t$ can be obtained as $\tilde{\pi}_t \pm z_{1-\alpha/2}SE(\tilde{\pi}_t)$, where $z_{1-\alpha/2}$ is the normal quartile at the significance level $\alpha/2$ and SE is the standard error.

Hong Kong ::: APPLICATIONS:
The first SARS case in Hong Kong was diagnosed on March 12, 2003, and 1,755 Hong Kong patients were hospitalized over the course of the epidemic (13). Of these cases, 1,456 patients were discharged and 299 patients died from SARS by July 31, 2003.

The top panels of figure 2 show real-time transition probabilities and associated 95 percent confidence intervals. The estimated death rate, $\tilde{p}_{1t}$, peaked in mid-March, decreased for approximately 2 weeks, and then increased again until mid-May. The estimated recovery rate $\tilde{p}_{2t}$, shown in the top right panel, rose in April, when the majority of SARS patients (mostly health-care workers) from a regional hospital (Prince of Wales Hospital) were discharged, and was then relatively steady until late June. The bottom left panel depicts the ratio of the two estimated transitional probabilities $\tilde{\theta}_t$, and the bottom right panel plots the estimated fatality rate at time t, $\tilde{\pi}_t$, along with the WHO fatality rate over the course of the epidemic for comparison.

The proposed estimate of the fatality rates was high at the beginning of the outbreak. This may reflect that the patients who were initially admitted were quite ill but, as the epidemic progressed, patients were admitted earlier in the progression of their disease. The estimated fatality rate stabilized at the beginning of April, and until mid-June it fluctuated between 15 percent and 25 percent. The stabilization of the estimated fatality rate could reflect a more mature approach in the diagnosis and treatment of SARS, although the efficacy of using high-dose corticosteroid and ribavirin to deal with lung parenchymal inflammation and the coronavirus itself is controversial (14–16). Alternatively, it could reflect a stationary distribution of the chain multinomial model. The WHO-adopted estimate increased gradually over the course of the epidemic and converged to the final overall fatality rate, which was approximately 17 percent.

Beijing ::: APPLICATIONS:
SARS data from Beijing have been available on the WHO website (http://www.who.int/csr) since April 21, 2003. By July 2, 2003, there had been 191 deaths among the 2,521 cases. Our estimates are plotted in the bottom right panel of figure 3. As in Hong Kong, the fatality rate was estimated to be high in the early stage of the epidemic. It then continuously improved over the course of the epidemic, with an overall fatality rate of approximately 8 percent. The reduction in the fatality rate is possibly related to the use of different treatments by the Chinese authorities, who also implemented effective measures to prevent infection in the community and in hospitals during the epidemic. In Beijing, a special hospital was built within 2 weeks of the start of the epidemic to treat SARS patients. Here, the WHO fatality rate overestimates the fatality rate in the latter part of the epidemic. The use of traditional Chinese medicine has also been suggested to have contributed to the low rate of fatality in Beijing, and glycyrrhizin, an active component of licorice root, has been shown to inhibit the replication of SARS-associated coronavirus (17).

SIMULATED DATA:
We simulated data under various scenarios to compare the proposed real-time fatality rate with the WHO fatality rate. In the simulation, we used the daily number of inpatients in the Hong Kong SARS epidemic as the basis for simulation for each day so that the infection process was not simulated. The numbers of deaths and recoveries for each day were then simulated according to the assumed multinomial probabilities under different scenarios. We chose bandwidths for the estimator by means of the empirical bias bandwidth selection method.

Scenario I assumes constant death and recovery rates on each day, with values of 0.01 and 0.1, respectively. This can be the situation if no effective treatment is available and/or the lethality of the disease does not change over time. Scenario II is similar to scenario I except for the appearance of a sudden increase in the death rate from 0.01 to 0.03 at 30 days after the start of the epidemic. This mirrors the Amoy Gardens outbreak in Hong Kong, where the strain of the SARS virus was found to be more lethal. In this scenario, it is assumed that the death rate then drops back to its initial level after a period of time. Scenario III assumes a consistently decreasing death rate and an increasing recovery rate. In this scenario, the fatality rate would decrease monotonically.

The left panels in figure 4 show the true death and recovery rates under the different scenarios. The right panels show the corresponding true daily fatality rate p1t /(p1t + p2t), the average over 100 simulations of the proposed real-time fatality rate, and the WHO fatality rate. In all cases, the mean of the proposed fatality rate estimated the true fatality rate quite well and captured the trend or sudden changes in the true fatality for the first and third scenarios. In the second scenario, the abrupt change was not totally captured, but this is to be expected.

In scenario I, the WHO fatality rate increased consistently during the first 60 days and then stabilized, so it was underestimating the true fatality rate in the earlier period of the epidemic. In scenario II, the WHO fatality rate could only gradually reflect the sudden increase in the true fatality rate, but it still underestimated this rate. In addition, the WHO fatality rate did not reflect the later decrease in the fatality rate. In scenario III, after an initial increase, the WHO fatality rate was relatively constant and underestimated the decrease in the fatality rate.

DISCUSSION:

We have shown that a real-time fatality rate estimated by means of kernel smoothing methods has major advantages over the traditional WHO estimate during the course of an epidemic. An iteratively reweighted kernel estimator

$$\tilde{p}_{it} = \frac{\sum_{s=1}^{T} K_b(t-x)\hat{p}_{is}w_{is}}{\sum_{s=1}^{T} K_b(t-x)w_{is}},$$

where

$$w_{is} = \frac{H_{s-1}}{\tilde{p}_{it}^{\ast}(1-\tilde{p}_{it}^{\ast})}$$

is the weight to be iterated until convergence, can also be used. This estimator should be more efficient, but the improvement is expected to be marginal. In general, the weight function, which puts emphasis on the information in different periods or characteristics contributed to the estimator, does not affect the estimator much in comparison with the choice of bandwidth.

Our simulation studies show that variations in clinical outcome measures—resulting from improved treatment modalities, for example—are not accurately detected by the WHO estimator. On the other hand, our method reflects the possible impact of such factors, including variation in clinical management and other exogenous factors over the course of the outbreak. The spread of the SARS epidemic in Hong Kong was similar to that of scenario I, in which the recovery and mortality rates are nearly constant. In addition, the situation in Beijing was similar to that of scenario III, in which constant improvement in rates of recovery and death was observed. However, as was shown in the simulation, the WHO estimate fails to reflect the time-varying fatality rate in all regions.

The WHO estimates are simply not suitable for monitoring the fatality rate over the course of an emerging epidemic. The present formulation is also different from traditional survival analysis, which adopts chronologic time on the individual level rather than calendar time on the population level and thus requires more detailed data. The proposed chain multinomial model of using the real-time case fatality rate should allow epidemiologists to monitor not only SARS but also other emerging infections and conditions in the future. It can provide a timely and accurate estimate of the number of fatalities inflicted by an epidemic, especially in the earlier period when decisions on health policy and treatment are most critical. It works satisfactorily if the durations of hospital stays for recoveries and deaths are not that different. Yip et al. (18) adopted a competing risk model to provide an estimate of the real-time fatality for SARS using a counting process approach. In the absence of data from controlled clinical trials on SARS treatment, it is interesting to speculate as to whether the different estimated fatality rates between Hong Kong and Beijing could be related to the different clinical treatment protocols used.