Lifetime Data Anal (2009) 15:316–329
DOI 10.1007/s10985-009-9119-3

Estimating the case fatality rate using a constant cure-death hazard ratio

Zheng Chen · Kohei Akazawa · Tsuyoshi Nakamura

Received: 6 June 2008 / Accepted: 5 May 2009 / Published online: 21 May 2009

Abstract The case fatality rate is an important indicator of the severity of a disease, and unbiased and accurate estimates of it during an outbreak are important in the study of epidemic diseases, including severe acute respiratory syndrome (SARS). In this paper, estimation methods are developed using a constant cure-death hazard ratio. A semiparametric model is presented, in which the cure-death hazard ratio is a parameter of interest, and a profile likelihood-based technique is proposed for estimating the case fatality rate. An extensive simulation was carried out to investigate the performance of this technique for small and medium sample sizes, using both summary and individual data. The results show that the performance depends on the model validity but is not heavily dependent on the sample size. The method was applied to summary SARS data obtained from Hong Kong and Singapore.

Keywords Case fatality rate · Competing risks · Constant cure-death hazard ratio · Profile likelihood · SARS

Z. Chen (✉)
Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou 510515, China
e-mail: zheng-chen@hotmail.com

K. Akazawa
Department of Medical Informatics, Niigata University Medical and Dental Hospital, Niigata 951-8520, Japan
e-mail: akazawa@med.niigata-u.ac.jp

T. Nakamura
Graduate School of Science and Technology, Nagasaki University, Nagasaki 852-8521, Japan
e-mail: naka@nagasaki-u.ac.jp

Abbreviations
SARS Severe acute respiratory syndrome
WHO World Health Organization

1 Introduction

Severe acute respiratory syndrome (SARS) emerged at the end of 2002, and of the 8437 cases reported prior to July, 2003, 813 individuals died (WHO 2003a). For a new epidemic of such a disease, the case fatality rate is of considerable interest and has been discussed by the World Health Organization (WHO 2003b), Donnelly et al. (2003), Chen and Nakamura (2004), Yip et al. (2005), Ghani et al. (2005), Yu et al. (2006), Jewell et al. (2007), Chang et al. (2007), among others.

The case fatality rate of a disease is defined as the proportion of cases, among those who develop a disease, who then proceed to die from the disease (Rothman 2002). This definition can only be used to compute the exact value of the case fatality rate at the end of an epidemic. Estimates calculated during an outbreak may be subject to bias because of future deaths that have not yet been counted. The estimation methods are classified according to whether individual (where individual survival times are included) or summary data are given.

For summary data (i.e., data collected daily over a given period of time, showing the total number of admissions, deaths, and cures) during an outbreak of SARS, the WHO (2003b) presented two traditional estimates: the ratio of deaths among all reported SARS cases, and the ratio of deaths among SARS cases with a known outcome (patients who are known to have recovered or died). Chen and Nakamura (2004) proposed a cure-death hazard plot, a scatter plot of the cumulative hazards for recovery versus death, based on a competing risks theory. The ultimate case fatality rate can then be estimated by extrapolation using a regression equation. On the other hand, Yip et al. (2005) defined a real-time SARS fatality rate as the probability of death, conditioned on a transition to either death or recovery, but they did not estimate the overall case fatality rate.

For individual data, Donnelly et al. (2003) assumed gamma distributions for the individual time-delay distributions, admission-to-death, and admission-to-discharge. In addition, they presented a nonparametric estimation technique based on the Kaplan-Meier curve. Ghani et al. (2005) and Jewell et al. (2007) applied the Kaplan-Meier curve to Hong Kong data collected in 2003, jointly considering two outcomes (death and recovery). According to the results of Ghani et al. (2005), the extensional Kaplan-Meier method gave better estimates than the parametric cure model. Chang et al. (2007) obtained an age-specific case fatality rate for the Taiwan SARS epidemic by applying a modified semiparametric cure model, previously used by Fine (1999), to data collected throughout the epidemic.

In this paper, a likelihood-based method is developed that is applicable to both summary data and individual data with covariates. The next section describes the profile likelihoods for analyzing the two types of data. In Sect. 3, numerical results are used to compare the performance of those likelihood methods with the techniques given in the literature. Section 4 describes summary SARS data for Hong Kong and Singapore.

This study deals with discrete failure time models, since the survival experiences of SARS patients were recorded daily and included a considerable number of ties for individual data.

2 Estimation methods

Let T be a random variable that measures the time elapsed from admission to the occurrence of a failure. Suppose that there are two competing types of failure, denoted by $J = 1$ and $J = 2$. In this paper, the failure type $J = 1$ will be referred to as *death* and $J = 2$ will be referred to as *cure*.

2.1 Summary data

Assume that failures occur at discrete times $t_1 < t_2 < \cdots < t_i < \cdots < t_s < \cdots < t_{\max}$, where t_s is the time when the study is carried out, and all patients experience failure no later than $t_{\max}$. Let $S(t) = P(T \geq t)$ denote the survivor function, $h_{ji} = h_j(t_i) = P(T = t_i, J = j | T \geq t_i)$ a type-j hazard at t_i, for $j = 1, 2$ and $i = 1, 2, \ldots,$ respectively. The time t_i will be referred to as Day i. Let π_1 and π_2 denote the case fatality rate and case cure rate, respectively, with $\pi_1 + \pi_2 = 1$. Define the cure-death hazard ratio at t_i as $\theta_i = h_{2i}/h_{1i}$. If the ratio is a constant $\theta_i \equiv \theta$ for $i = 1, 2, \ldots,$ then the case fatality rate is given by $\pi_1 = (1 + \theta)^{-1}$.

Let $n_i, d_i, c_i, N_i, D_i,$ and C_i denote the number of new admissions, deaths, and cures, and the cumulative number of admissions, deaths, and cures at time t_i $(i = 1, 2, \ldots, s)$, respectively. It is assumed that no patient dies or is cured at the same time, t_i, that he/she is admitted to a hospital. Then the number of patients at risk at time t_i is $a_i = N_{i-1} - D_{i-1} - C_{i-1}$. Since the probability of no failures at time t_i is $1 - h_{1i} - h_{2i}$, the contribution of the observations c_i and d_i given a_i to the log-likelihood (Betensky and Schoenfeld 2001) is

$$l_i = d_i \log(h_{1i}) + c_i \log(h_{2i}) + (a_i - d_i - c_i) \log(1 - h_{1i} - h_{2i}).$$

The martingale argument of counting process theory (Andersen et al. 1993) then leads to an overall log-likelihood of $l = \sum_{i=1}^{s} l_i$ for the hazards, h_{1i} and $h_{2i}, i = 1, 2, \ldots, s$.

Let us assume that the cure-death hazard ratio is time-independent, $h_{2i}/h_{1i} \equiv \theta$. Then the log-likelihood for θ and the h_{1i} is

$$l = \sum_{i=1}^{s} \{d_i \log(h_{1i}) + c_i \log(\theta h_{1i}) + (a_i - d_i - c_i) \log(1 - h_{1i} - \theta h_{1i})\}$$

Solving the equation $\partial l / \partial h_{1i} = 0$, we have $\hat{h}_{1i} = (d_i + c_i)/[a_i(1 + \theta)]$. Assigning $\hat{h}_{1i}$ to h_{1i} leads to the profile log-likelihood (Murphy and Van der Vaart 2000) for θ

$$pl = \sum_{i=1}^{s} \left\{ d_i \log \frac{d_i + c_i}{a_i(1+\theta)} + c_i \log \frac{\theta(d_i + c_i)}{a_i(1+\theta)} + (a_i - d_i - c_i) \log \left(1 - \frac{d_i + c_i}{a_i}\right) \right\}.$$

The maximum profile likelihood estimate and an asymptotic variance estimate based on the profile information matrix (Tsodikov et al. 2007) are given by

$$\hat{\theta} = \frac{\sum_{i=1}^{s} c_i}{\sum_{i=1}^{s} d_i} \quad \text{and} \quad avar(\hat{\theta}) = \frac{\sum_{i=1}^{s} c_i \sum_{i=1}^{s} (d_i + c_i)}{\left(\sum_{i=1}^{s} d_i\right)^3}.$$

The estimate of the case fatality rate and an asymptotic variance estimate based on the delta method are then given by

$$\hat{\pi}_1 = 1/(1+\hat{\theta}) \quad \text{and} \quad avar(\hat{\pi}_1) = avar(\hat{\theta}) \left(1+\hat{\theta}\right)^{-4}.$$

2.2 Extension to individual data with covariates

If $\mathbf{Z}$ denotes a $1 \times w$ vector of covariates for an individual, the survival function with covariate $\mathbf{Z}$ will be denoted by $S(t, \mathbf{Z}) = \Pr(T \geq t|\mathbf{Z})$. The type-$j$ hazard function is

$$\lambda_j(t, \mathbf{Z}) = \Pr(T = t, J = j | T \geq t, \mathbf{Z}), \quad j = 1, 2, \tag{1}$$

and the type-j frequency function is $f_j(t, \mathbf{Z}) = \Pr(T = t, J = j|\mathbf{Z})$, for $j = 1, 2$. Let $\pi_1(\mathbf{Z})$ and $\pi_2(\mathbf{Z})$denote the covariate specific case fatality rate and case cure rate, respectively. The cure-death hazard ratio given $\mathbf{Z}$ at t is $\theta(t, \mathbf{Z}) = \lambda_2(t, \mathbf{Z})/\lambda_1(t, \mathbf{Z})$. If the ratio is time-independent over the study period, say $\theta(t, \mathbf{Z}) \equiv \theta(\mathbf{Z})$, then, $\pi_1(\mathbf{Z}) = (1 + \theta(\mathbf{Z}))^{-1}$.

Lunn and McNeil (1995) assume a proportional hazards model for the type-specific hazard functions in Eq. 1,

$$\lambda_j(t, \mathbf{Z}) = \lambda_{j0}(t) \exp\left[\boldsymbol{\beta}_j^T \mathbf{Z}\right], \quad j = 1, 2,$$

where $\lambda_{j0}(t)$ is a baseline hazard, and $\boldsymbol{\beta}_j$ is a $1 \times w$ vector of regression coefficients, for $j = 1, 2$. The sub-survivor function and sub-density function are, respectively,

$$S_j(t, \mathbf{Z}) = \exp\left(-\sum_{t_i \leq t} \lambda_j(t_i, \mathbf{Z})\right) \quad \text{and} \quad f_j(t, \mathbf{Z}) = \lambda_j(t, \mathbf{Z}) S(t, \mathbf{Z}),\ j = 1, 2,$$

where $S(t, \mathbf{Z}) = S_1(t, \mathbf{Z}) S_2(t, \mathbf{Z})$ is the overall survivor function. Correspondingly, the cure-death hazard ratio is written as $\theta(t, \mathbf{Z}) = \theta(t) \exp[(\boldsymbol{\beta}_2^T - \boldsymbol{\beta}_1^T)\mathbf{Z}]$, where $\theta(t) = \lambda_{20}(t)/\lambda_{10}(t)$ is a baseline cure-death hazard ratio. Hereafter, we assume $\theta(t)$ is a constant θ_0 over the study period, and therefore,

$$\lambda_{20}(t) = \theta_0 \lambda_{10}(t) \quad \text{and} \quad \theta(\mathbf{Z}) = \theta_0 \exp\left[\left(\boldsymbol{\beta}_2^T - \boldsymbol{\beta}_1^T\right)\mathbf{Z}\right].$$

Assume that $(\tau_k, \delta_k, j_k, \mathbf{Z}_k)$ is observed for each subject k, $k = 1, 2, \ldots, m$ (m is the number of subjects), where τ_k denotes the time elapsed until either a failure or a

censoring occurs, j_k is an indicator ($j_k = 1$ for death, 2 for cure, and 0 for censoring), and $\mathbf{Z}_k$ is a covariate vector. Define $\delta_{1k} = \mathrm{I}[j_k = 1]$, $\delta_{2k} = \mathrm{I}[j_k = 2]$, where $\mathrm{I}[q]$ is 1 if q is true, and 0 otherwise. The likelihood is then given by

$$L = \prod_{k=1}^{m} L_k = \prod_{k=1}^{m} \left\{ \prod_{j=1}^{2} \lambda_j(\tau_k, \mathbf{Z}_k)^{\delta_{jk}} S_1(\tau_k, \mathbf{Z}_k) S_2(\tau_k, \mathbf{Z}_k) \right\}.$$

For notational simplicity, set $\beta_0 = \log\theta_0$ and $\boldsymbol{\beta}^T = (\beta_0, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$, and change the order of the summations. The profile log-likelihood $pl(\boldsymbol{\beta})$ is then the sum of

$$pl_i(\boldsymbol{\beta}) = -\Delta_i \log \left\{ \sum_{k \in R_i} \exp\left(\boldsymbol{\beta}_1^T \mathbf{Z}_k\right) + \sum_{k \in R_i} \exp\left(\beta_0 + \boldsymbol{\beta}_2^T \mathbf{Z}_k\right) \right\} + \sum_{k \in R_{1i}} \boldsymbol{\beta}_1^T \mathbf{Z}_k + \Delta_{2i}\beta_0 + \sum_{k \in R_{2i}} \boldsymbol{\beta}_2^T \mathbf{Z}_k$$

for $i = 1, \ldots, s$, where $R_{ji} = \{k | \tau_k = t_i, \delta_{jk} = 1\}$ denotes the set of individuals who fail at t_i with type-j failure, and R_i those at risk at t_i, Δ_{1i} and Δ_{2i} are the number of deaths and cures, respectively, at t_i and $\Delta_i = (\Delta_{1i} + \Delta_{2_i})$ (Appendix A). The maximum profile likelihood estimate $\hat{\boldsymbol{\beta}}^T = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T)$ and the asymptotic variance estimate $avar(\hat{\boldsymbol{\beta}})$ are then obtained by maximizing $pl(\boldsymbol{\beta})$ (Appendix B).

The estimate for the constant cure-death hazard ratio with covariate $\mathbf{Z}$ is,

$$\hat{\theta}(\mathbf{Z}) = \hat{\theta}_{\mathbf{0}} \exp\left[\left(\hat{\boldsymbol{\beta}}_2^T - \hat{\boldsymbol{\beta}}_1^T\right)\mathbf{Z}\right] = \exp\left[\hat{\beta}_0 - \hat{\boldsymbol{\beta}}_1^T\mathbf{Z} + \hat{\boldsymbol{\beta}}_2^T\mathbf{Z}\right].$$

Since $\log\hat{\theta}(\mathbf{Z}) = (1, -\mathbf{Z}^T, \mathbf{Z}^T)\hat{\boldsymbol{\beta}}$, its asymptotic variance estimate is given by

$$avar[\log\hat{\theta}(\mathbf{Z})] = (1, -\mathbf{Z}^T, \mathbf{Z}^T) avar(\hat{\boldsymbol{\beta}}) (1, -\mathbf{Z}^T, \mathbf{Z}^T)^T.$$

Then the asymptotic variance estimates (obtained from the delta method) for $\hat{\theta}(\mathbf{Z})$ and $\hat{\pi}_1(\mathbf{Z}) = [1 + \hat{\theta}(\mathbf{Z})]^{-1}$ are, respectively,

$$avar[\hat{\theta}(\mathbf{Z})] = avar[\log\hat{\theta}(\mathbf{Z})]\hat{\theta}(\mathbf{Z})^2$$

and

$$avar[\hat{\pi}_1(\mathbf{Z})] = avar[\hat{\theta}(\mathbf{Z})][1 + \hat{\theta}(\mathbf{Z})]^{-4}.$$

If $\mathbf{Z} \equiv 0$ or $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, then $\theta(\mathbf{Z})$ and $\pi_1(\mathbf{Z})$ are θ_0 and π_1, respectively. In this simple case, $pl(\boldsymbol{\beta})$ reduces to $pl(\theta_0) = \sum_{i=1}^{s} \left\{-(\Delta_{1i} + \Delta_{2_i})\log(1 + \theta_0) + \Delta_{2_i}\log\theta_0\right\}$ and the maximum profile likelihood estimate and asymptotic variance estimate are, respectively,

$$\hat{\theta}_0 = \frac{\sum_{i=1}^{s} \Delta_{2i}}{\sum_{i=1}^{s} \Delta_{1i}} \quad \text{and} \quad avar(\hat{\theta}_0) = \frac{\sum_{i=1}^{s} \Delta_{2i} \sum_{i=1}^{s} (\Delta_{1i} + \Delta_{2i})}{\left(\sum_{i=1}^{s} \Delta_{1i}\right)^3}.$$

3 Numerical results

3.1 Summary data

The first subsection assesses the performance of the profile likelihood estimates for summary data, as presented in Sect. 2.1. Comparisons between the empirical variance and the asymptotic variance estimate are of particular interest. A uniformly distributed random number U in (0, 1) is generated daily for each subject at risk. The outcomes are then death, cure, or still at risk at time $(i + 1)$, depending on whether $U \leq h_{1i}$, $h_{1i} < U \leq h_{1i} + h_{2i}$, or $h_{1i} + h_{2i} < U$. Parameter values and sample sizes are chosen based on actual SARS data described in Sect. 4, and 10000 independent estimates are generated.

Data are simulated under 3 scenarios. The initial number of subjects is $a_1 = 200$ in scenario I, $a_1 = 1500$ in scenario II, and $a_1 = 100$ in scenario III. In both scenarios I and II, there are no new admissions on any day ($n_i = 0$ for all i). In scenario III, the number of new daily admissions is 25 on Days 1–15, 50 on Days 16–35, and 15 on Days 36–45. The number of subjects at risk on Day i is $a_i = a_{i-1} + n_{i-1} - d_{i-1} - c_{i-1}$ for $i = 2, 3, \ldots, s$. In each scenario, the data are evaluated on Day s for $s = 20$, $s = 30$, and $s = 40$. In all scenarios, the death hazard rate h_{1_i} is drawn from a normal distribution with a mean of 0.01 and a variance of either 1×10^{-5} or 2.5×10^{-5}. The cure-death hazard ratio is $\theta = 4.5$ and the cure hazard rate is $h_{2_i} = \theta h_{1_i}$. If the generated value of h_{1_i} is negative, it is discarded and a new value is generated.

Table 1 shows the results. When $a_1 = 200$ and $n_i = 0$, the bias in $\hat{\theta}$ is between 1.9 and 3.4%. Otherwise, for scenarios (II) and (III), the biases in the estimates are less than 1.5%. As for the variances, the greatest bias in $avar(\hat{\theta})^{1/2}$ is 2.9%, occurring on Day 20, when $a_1 = 100$ and the variance of h_1 is 2.5×10^{-5}. These results indicate the approximate unbiasedness of $avar(\hat{\theta})$ for the sample sizes actually observed in the Singapore data.

3.2 Individual data without covariates

This section deals with the estimates for individual data without covariates. The estimates obtained from our method are compared to those obtained from a parametric cure model that assumes a gamma distribution. The sample size is either 1500 or 200, and 1000 independent estimates are generated at each of three different points in time, Day $(s_m - 7)$, s_m and $(s_m + 7)$, where Day s_m denotes the average time from admission to death.

The first sample is generated following the cure model of Donnelly et al. (2003). Gamma or Weibull distributions are used to generate times from admission to death and from admission to cure. The following parameter values are chosen based on Donnelly et al. (2003). The case fatality rate is $\pi_1 = 18.18\%$, the mean and variance

Table 1 Estimates of the constant cure-death hazard ratio θ when the hazards are normal random variables

	Variance of h_1	Day s	$\hat{\theta}$	SD($\hat{\theta}$)	$avar(\hat{\theta})^{1/2}$
I	1×10^{-5}	20	4.66	1.14	1.13
		30	4.63	1.00	1.00
		40	4.61	0.94	0.95
	2.5×10^{-5}	20	4.60	1.07	1.09
		30	4.59	0.99	0.98
		40	4.59	0.94	0.93
II	1×10^{-5}	20	4.51	0.37	0.37
		30	4.50	0.33	0.34
		40	4.50	0.31	0.32
	2.5×10^{-5}	20	4.55	0.37	0.37
		30	4.55	0.34	0.34
		40	4.55	0.33	0.32
III	1×10^{-5}	20	4.56	0.72	0.74
		30	4.52	0.50	0.50
		40	4.52	0.40	0.39
	2.5×10^{-5}	20	4.56	0.76	0.74
		30	4.53	0.50	0.50
		40	4.52	0.39	0.38

True values are $\theta = 4.5$. (I) $a(1) = 200$ with $n_i = 0$; (II) $a(1) = 1500$ with $n_i = 0$; (III) $a(1) = 100$ with varied n_i

of the admission-to-death distribution are 36 and 573, respectively, and the mean and variance of the admission-to-cure distribution are 23 and 62, respectively. The simulation results are shown in Table 2. Our method yields more biased estimates, while the parametric cure model assuming gamma distributions yields approximately unbiased estimates on Day 44.

The second sample is generated in accordance with a nonparametric competing risks model. For each individual, a uniform random number $U_t(0, 1)$ is generated at each t. The outcome at t is then death if $U_t \leq \lambda_1(t)$, cure if $\lambda_1(t) < U_t \leq \lambda_1(t) + \lambda_2(t)$, and no failure otherwise. An individual is at risk at time $(t + 1)$ if the outcome is no failure at t. In this simulation, $\theta \equiv 4.5$, $\pi_1 = 18.18\%$, and $\lambda_1(t)$ is assumed to be $0.001 \times t/10$ (monotone increasing) or normally distributed with either mean 0.01 and variance 2.5×10^{-5} or mean 0.005 and variance 5×10^{-6}. Table 3 shows the results. As expected, our estimates are approximately unbiased in all cases, whereas the parametric cure model is biased.

3.3 Individual data with covariates

This section examines the performance of our method for individual data with a covariate Z. The parameter values are chosen on the basis of observations by

Table 2 Comparison of estimates $\hat{\pi}_1(\%)$ when survival times follow a parametric distribution

$T^{\dagger}$	Sample size	Day s	Profile likelihood			Cure model assuming γ distribution			
			$\hat{\pi}_1$	SD($\hat{\pi}_1$)	$avar(\hat{\pi}_1)^{1/2}$	$\hat{\pi}_1$	SD($\hat{\pi}_1$)	$avar(\hat{\pi}_1)^{1/2}$	$f^{\dagger\dagger}$
Gamma distribution	200	30	11.48	2.607	2.577	13.78	4.653	5.063	6
		37	12.29	2.483	2.463	16.48	3.730	3.619	2
		44	13.51	2.524	2.493	17.66	2.925	2.903	2
	1500	30	11.42	0.957	0.942	13.48	1.644	1.896	6
		37	12.27	0.930	0.901	16.56	1.435	1.373	5
		44	13.50	0.935	0.912	17.75	1.113	1.054	1
Weibull distribution	200	30	10.62	2.503	2.423	11.18	2.988	3.364	5
		37	11.67	2.465	2.378	13.55	3.032	3.106	19
		44	13.08	2.509	2.451	16.37	2.874	2.876	7
	1500	30	10.69	0.895	0.890	11.32	1.096	1.093	27
		37	11.71	0.879	0.872	13.55	1.102	1.144	14
		44	13.12	0.904	0.898	16.44	1.038	1.047	29

True $\pi_1 = 18.18$

† Distribution assumed for survival time

†† f denotes the number of iterations deleted due to encountering a singular information matrix in the course of the Newton–Raphson algorithm

Leung et al. (2004). A covariate $Z = 0$ corresponds to an age younger than 60, and $Z = 1$ to ages 60 and older. Leung et al. reported that $\pi_1(0)$ and $\pi_1(1)$ are 7% and 54.5%, respectively, while the percentage of patients with $Z = 0$ and $Z = 1$ are approximately 80% and 20%, respectively. Next we generate $\lambda_1(t, 0)$ to follow the normal distribution with mean 0.001 and standard deviation 0.0005, so that $\theta(0)(=\lambda_2(t, 0)/\lambda_1(t, 0))$ and $\theta(1)(=\lambda_2(t, 1)/\lambda_1(t, 1))$ are approximately 13.3 and 0.8, or $\pi_1(0) = 7\%$ and $\pi_1(1) = 54.5\%$, respectively. A uniform random number U_t is generated at each time for each individual at risk. The outcome at t is death if $U_t \leq \lambda_1(t, Z)$, cure if $\lambda_1(t, Z) < U_t \leq \lambda_1(t, Z) + \lambda_2(t, Z)$, and at risk at time $(t + 1)$ otherwise. The sample sizes are 1500 and 200, and 1000 independent estimates are generated for each parameter combination.

Table 4 shows the results. The bias in $\hat{\pi}_1(0)$ is less than 3% on Day 30 when the sample size is 1500. As for the standard deviation, the bias for $Z = 0$ and $Z = 1$ are roughly 3% and 7%, respectively, when $n = 200$. Hence the estimates are approximately unbiased with valid variance estimates, except for $n = 200$ and $Z = 1$, with estimated variances slightly smaller than the observed ones.

4 Application to summary SARS data

The method of Sect. 2.1 is applied to summary SARS data from Hong Kong and Singapore. By the end of the epidemic, there were 298 fatalities among the 1755 cases reported in Hong Kong, and 32 fatalities among the 206 cases in Singapore (WHO 2003a). The data used in this study were collected over the Internet from information posted by the Department of Health of the Hong Kong Special Administrative Region,

Table 3 Comparison of estimates $\hat{\pi}_1(\%)$ when $\lambda_1(t)$ is increasing with t or follows a normal distribution

$\lambda_1(t)$	Sample size	Day[§] s	Profile likelihood			Cure model assuming γ distribution			
			$\hat{\pi}_1$	SD($\hat{\pi}_1$)	$avar(\hat{\pi}_1)^{1/2}$	$\hat{\pi}_1$	SD($\hat{\pi}_1$)	$avar(\hat{\pi}_1)^{1/2}$	f
$0.001 \times t$ / 10	200	46	18.07	4.164	4.045	19.23	10.65	18.84	31
		53	18.17	3.811	3.676	18.93	7.688	11.07	23
		60	18.09	3.530	3.402	18.35	5.815	7.306	11
	1500	46	18.20	1.531	1.484	18.23	3.222	5.160	110
		53	18.20	1.374	1.345	18.18	2.495	3.412	21
		60	18.20	1.268	1.246	18.18	1.940	2.427	17
Normal distribution N(0.01,0.000025)	200	10	18.63	4.067	4.158	26.16	18.43	15.33	27
		17	18.51	3.455	3.462	23.48	11.59	8.056	19
		23	18.49	3.232	3.185	21.96	8.139	6.037	25
	1500	11	18.16	1.474	1.455	26.07	15.41	6.719	180
		18	18.15	1.223	1.237	22.91	9.439	3.634	234
		24	18.16	1.141	1.148	21.35	6.285	2.602	230
Normal distribution N(0.005,0.000005)	200	28	18.67	3.728	3.721	25.18	14.04	16.41	32
		35	18.59	3.434	3.461	23.39	10.84	9.538	26
		42	18.50	3.265	3.282	21.58	8.254	8.059	29
	1500	29	18.18	1.299	1.333	23.64	11.35	6.267	111
		36	18.22	1.247	1.248	22.24	8.549	4.984	112
		43	18.20	1.179	1.188	20.77	6.320	5.795	107

True $\pi_1 = 18.18$

[§] Three figures of Day s are $s_m - 7$, s_m and $s_m + 7$ with $s_m =$ average time from admission to death as described in Sect. 3.2

Table 4 Estimates of covariate-specific case fatality rate $\pi_1(Z)(\%)$

Sample size	Day s	Age$< 60(Z = 0)$			Age$\geq 60(Z = 1)$		
		$\hat{\pi}_1(0)$	SD$[\hat{\pi}_1(0)]$	$avar[\hat{\pi}_1(0)]^{1/2}$	$\hat{\pi}_1(1)$	SD$[\hat{\pi}_1(1)]$	$avar[\hat{\pi}_1(1)]^{1/2}$
200	30	6.98	3.46	3.36	54.56	13.23	12.40
	40	6.95	3.10	3.00	54.59	11.99	11.15
	50	6.93	2.86	2.77	54.42	11.16	10.37
1500	30	7.08	1.27	1.24	54.66	4.61	4.54
	40	7.04	1.14	1.11	54.59	4.17	4.08
	50	7.03	1.04	1.02	54.56	3.81	3.79

True $\pi_1(0) = 7\%$ and $\pi_1(1) = 54.5\%$

the People's Republic of China, the Ministry of Health of Singapore, and the World Health Organization Cumulative Number of Reported Probable Cases of SARS.

In Hong Kong, the first complete report on SARS patients was released via the Internet on 19 March, 2003, with the first patient having been seen on 15 February, 2003.

Table 5 Estimates of the case fatality rate $\hat{\pi}_1(\%)$ of summary SARS data by area

	Study point	Profile likelihood method		End of epidemic
		$\hat{\pi}_1$	(95% CI)	π_1
Hong Kong	23 Apr	16.48	(13.58–19.38)	17.10
	7 May	17.03	(14.89–19.17)	
	21 May	16.98	(15.08–18.88)	
	4 Jun	17.35	(15.51–19.19)	
Singapore	23 Apr	13.18	(7.34–19.02)	15.53
	7 May	15.43	(10.08–20.78)	
	21 May	15.43	(10.29–20.57)	
	4 Jun	15.98	(10.83–21.13)	
Beijing	21 Apr–5 May	49.13*	(41.68–56.58)	7.62
	6 May–19 May	17.09*	(12.64–21.54)	
	20 May–29 May	4.30*	(2.69–5.91)	
	30 May–13 Jun	1.07*	(0.44–1.70)	

*Phase-specific fatality rate. Three change time points are 5, 19 and 29 May 2003 (Chen and Nakamura 2004)

The corresponding dates for Singapore are 14 March and 25 February, 2003. Although the data were updated daily, no information on individual patients was included. Six summary figures were reported for each day t_i, namely n_i, d_i, c_i, N_i, D_i, and C_i. For Hong Kong, t_1 was 19 March, 2003, and for Singapore, t_1 was 14 March, 2003. The case fatality rates were estimated at four different points in time (23 April, 7 May, 21 May, and 4 June, 2003), using only data collected from the beginning to each point in time. Table 5 shows that the estimates are approximately equal to the final fatality rate, π_1, in the last column, and all 95% confidence intervals $\hat{\pi}_1 \pm 1.96 \times avar(\hat{\pi}_1)^{1/2}$ include π_1.

5 Discussion

As one referee pointed out, the constant cure-death hazard ratio assumption simplifies the estimation problem. Ghani et al. (2005) declared, "To obtain an estimate, we must make an assumption about the pattern of deaths and discharges beyond the point of observation. A sensible assumption is that the remaining outcomes occur with the same relative probabilities as observed up to the time of analysis." In general, their theory seems to apply to any prediction model, although Jewell et al. (2007) do not explicitly mention any assumption in order for their estimate to be unbiased. In fact, the assumption holds for the SARS data for Hong Kong and Singapore, which is to say, the cure-death hazard ratio was constant throughout the study period (Chen and Nakamura 2004). This observation prompted us to develop a method for estimating the case fatality rate, assuming that the cure-death hazard ratio is constant. The assumption provides a profile likelihood that is applicable to summary data as well as to individual data.

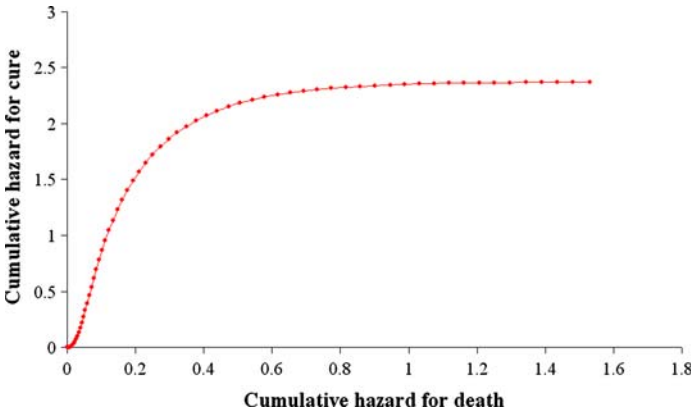

Fig. 1 Cure-death hazard plot for data assuming gamma distribution in Table 2

For individual data, the covariate-specific case fatality rate $\pi_1(\mathbf{Z})$ is written as

$$\pi_1(\mathbf{Z}) = \{1 + \theta(\mathbf{Z})\}^{-1} = \{1 + \exp[\beta_0 - \boldsymbol{\beta}_1^T \mathbf{Z} + \boldsymbol{\beta}_2^T \mathbf{Z}]\}^{-1}.$$

This simple form makes it possible to estimate the covariate-specific case fatality rate with a valid variance estimate. Lunn and McNeil (1995) suggested a method for estimating β_0, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ using the partial likelihood for the competing risks model. They assumed $\lambda_{20}(t) = \lambda_{10}(t) \exp(\beta_0)$ as in the present study, but they also made the assumption that there are no tied failures.

The numerical results indicate that the cure model assuming gamma distributions yields approximately unbiased estimates only when the failure times actually follow gamma distributions and the estimation is performed after the peak of the epidemic (Table 2). Table 2 indicates that the profile likelihood method yields estimates that are biased toward zero. The reason for this is understood by examining Fig. 1, which shows the cure-death hazard plot for the gamma distributions used in the simulations. The cure-death hazards ratio is high during the early stage and gradually decreases thereafter. Thus, the profile likelihood estimate assuming a constant hazards ratio underestimated the true value during the early stage. To avoid such difficulties, the cure-death hazard plot is examined before applying the profile likelihood method. A comprehensive testing method for the constant hazards ratio between the competing risks is discussed in Lam et al. (2008).

As regards the present application, the 95% confidence interval estimates made as early as April 23 are still accurate for Hong Kong and Singapore, as expected from the results shown in Table 5. Nevertheless, the application was limited to summary data only, since the individual SARS data are not currently available to the authors. We hope that the method developed here will be applied to individual SARS data in certain institutions, and that the accuracy of the estimates of the age-specific fatality rate at different points in time will be compared with those of the cure models (Donnelly et al. 2003; Chang et al. 2007).

Chen and Nakamura (2004) indicated a phase-specific fatality rate for Beijing SARS data, where the cure-death hazard ratio was not constant over the period of study. On account of this, we extended the estimation method of Sect. 2.1 to a phase-specific model to describe how the fatality rate changes as time passes. The results (Table 5) show a case fatality rate that decreased drastically over time.

Appendix

A. Profile likelihood for individual data with covariates in Sect. 2.2

The likelihood is

$$
\begin{aligned}
L &= \prod_{k=1}^{m} L_k = \prod_{k=1}^{m} \left\{ \prod_{j=1}^{2} \lambda_j(\tau_k, \mathbf{Z}_k)^{\delta_{jk}} S_1(\tau_k, \mathbf{Z}_k) S_2(\tau_k, \mathbf{Z}_k) \right\} \\
&= \prod_{k=1}^{m} \left(\prod_{j=1}^{2} \lambda_j(\tau_k, \mathbf{Z}_k)^{\delta_{jk}} \prod_{j=1}^{2} \exp\left\{ -\sum_{u \le \tau_k} \lambda_j(u, \mathbf{Z}_k) \right\} \right) \\
&= \prod_{k=1}^{m} \left(\prod_{j=1}^{2} \left\{ \lambda_{j0}(\tau_k) \exp[\boldsymbol{\beta}_j^T \mathbf{Z}_k] \right\}^{\delta_{jk}} \prod_{j=1}^{2} \exp\left\{ -\sum_{u \le \tau_k} \lambda_{j0}(u) \exp[\boldsymbol{\beta}_j^T \mathbf{Z}_k] \right\} \right).
\end{aligned}
$$

Changing the order of the summations, we obtain

$$
\begin{aligned}
L &= \prod_{i=1}^{s} \left(\prod_{j=1}^{2} \left\{ \lambda_{j0}(t_i)^{\Delta_{ji}} \exp\left[\sum_{k \in R_{ji}} \boldsymbol{\beta}_j^T \mathbf{Z}_k \right] \right\} \prod_{j=1}^{2} \exp\left\{ -\sum_{k \in R_i} \lambda_{j0}(t_i) \exp[\boldsymbol{\beta}_j^T \mathbf{Z}_k] \right\} \right) \\
&= \prod_{i=1}^{s} \left(\prod_{j=1}^{2} \lambda_{j0}(t_i)^{\Delta_{ji}} \exp\left[\sum_{k \in R_{ji}} \boldsymbol{\beta}_j^T \mathbf{Z}_k \right] \prod_{j=1}^{2} \exp\left\{ -\lambda_{j0}(t_i) \sum_{k \in R_i} \exp[\boldsymbol{\beta}_j^T \mathbf{Z}_k] \right\} \right),
\end{aligned}
$$

where $R_{ji} = \{k | \tau_k = t_i, \delta_{jk} = 1\}$ denotes the set of individuals who fail at time t_i with type-j failure, and R_i those at risk at time t_i. Writing $\lambda_{10i} = \lambda_{10}(t_i)$,

$$
\begin{aligned}
\log L_i = \Delta_i \log \lambda_{10i} + \sum_{k \in R_{1i}} \beta_1^T \mathbf{Z}_k + \Delta_{2i} \log \theta_0 + \sum_{k \in R_{2i}} \boldsymbol{\beta}_2^T \mathbf{Z}_k \\
-\lambda_{10_i} \sum_{k \in R_i} \left(e^{\boldsymbol{\beta}_1^T \mathbf{Z}_k} + \theta_0 e^{\beta_2^T \mathbf{Z}_k} \right).
\end{aligned}
$$

The log-likelihood is $l = \sum_{i=1}^{s} \log L_i$. Solving the equation $\partial l / \partial \lambda_{10i} = 0$ yields

$$
\hat{\lambda}_{10i} = \frac{\Delta_{1i} + \Delta_{2i}}{\sum_{k \in R_i} \left[\exp(\boldsymbol{\beta}_1^T \mathbf{Z}_k) + \theta_0 \exp(\boldsymbol{\beta}_2^T \mathbf{Z}_k) \right]}.
$$

Let $pl_i(\boldsymbol{\beta})$ denote $\log L_i$ with λ_{10i} replaced by $\hat{\lambda}_{10i}$ for all i. The profile log-likelihood $pl(\boldsymbol{\beta}) = \sum_{i=1}^{s} pl_i(\boldsymbol{\beta})$ then behaves like an ordinary log-likelihood (Murphy and Van der Vaart 2000). Ignoring a constant, we have

$$pl_i(\boldsymbol{\beta}) = -\Delta_i \log \left\{ \sum_{k \in R_i} \exp\left(\boldsymbol{\beta}_1^T \mathbf{Z}_k\right) + \sum_{k \in R_i} \exp\left(\beta_0 + \boldsymbol{\beta}_2^T \mathbf{Z}_k\right) \right\} + \sum_{k \in R_{1i}} \boldsymbol{\beta}_1^T \mathbf{Z}_k + \Delta_{2i} \beta_0 + \sum_{k \in R_{2i}} \boldsymbol{\beta}_2^T \mathbf{Z}_k$$

B. The asymptotic variance estimate $avar(\hat{\boldsymbol{\beta}})$ for $\hat{\boldsymbol{\beta}}$

Set $S_{1i} = \sum_{k \in R_i} \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_k)$, $S_{2i} = \sum_{k \in R_i} \exp(\beta_0 + \boldsymbol{\beta}_2^T \mathbf{Z}_k)$, and $S_i = S_{1i} + S_{2i}$. Then

$$pl_i(\boldsymbol{\beta}) = -\Delta_i \log S_i + \sum_{k \in R_{1i}} \boldsymbol{\beta}_1^T \mathbf{Z}_k + \Delta_{2i} \beta_0 + \sum_{k \in R_{2i}} \boldsymbol{\beta}_2^T \mathbf{Z}_k.$$

Moreover, set $S_{1i}^* = \sum_{k \in R_i} \mathbf{Z}_k \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_k)$, $S_{2i}^* = \sum_{k \in R_i} \mathbf{Z}_k \exp(\beta_0 + \boldsymbol{\beta}_2^T \mathbf{Z}_k)$, $S_{1i}^{**} = \sum_{k \in R_i} \mathbf{Z}_k \mathbf{Z}_k^T \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_k)$ and $S_{2i}^{**} = \sum_{k \in R_i} \mathbf{Z}_k \mathbf{Z}_k^T \exp(\beta_0 + \boldsymbol{\beta}_2^T \mathbf{Z}_k)$. The components of the score vector are then obtained from their first derivatives:

$$\frac{\partial pl_i}{\partial \beta_0} = -\Delta_i S_i^{-1} S_{2i} + \Delta_{2i}, \quad \frac{\partial pl_i}{\partial \boldsymbol{\beta}_1} = -\Delta_i S_i^{-1} S_{1i}^* + \sum_{k \in R_{i1}} \mathbf{Z}_k,$$

$$\frac{\partial pl_i}{\partial \boldsymbol{\beta}_2} = -\Delta_i S_i^{-1} S_{2i}^* + \sum_{k \in R_{i2}} \mathbf{Z}_k.$$

Finally, $avar(\hat{\boldsymbol{\beta}})$ is obtained from their second derivatives (Tsodikov et al. 2007).

Acknowledgements The authors would like to thank the referees and the associate editor for their insightful comments and suggestions, which have greatly improved the paper.

References

Andersen PK, Borgan Ø, Gill RD et al (1993) Statistical models based on counting processes. Springer, New York

Betensky RA, Schoenfeld DA (2001) Nonparametric estimation in a cure model with random cure times. Biometrics 57:282–286. doi:10.1111/j.0006-341X.2001.00282.x

Chang IS, Hsiung CA, Wen CC et al (2007) Non-parametric maximum-likelihood estimation in a semiparametric mixture model for competing-risks data. Scand J Stat 34:870–895

Chen Z, Nakamura T (2004) Statistical evidence for the usefulness of Chinese medicine in the treatment of SARS. Phytother Res 18:592–594. doi:10.1002/ptr.1485

Donnelly CA, Ghani AC, Leung GM et al (2003) Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 361:1761–1766. doi:10.1016/S0140-6736(03)13410-1

Fine JP (1999) Analyzing competing risks data with transformation models. J R Stat Soc Ser B Stat Methodol 61:817–830. doi:10.1111/1467-9868.00204

Ghani AC, Donnelly CA, Cox DR et al (2005) Methods for estimating the case fatality ratio for a novel, emerging infectious disease. Am J Epidemiol 162:479–486. doi:10.1093/aje/kwi230

Jewell NP, Lei XD, Ghani AC et al (2007) Non-parametric estimation of the case fatality ratio with competing risks data: an application to severe acute respiratory syndrome (SARS). Stat Med 26:1982–1998. doi:10.1002/sim.2691

Lam KF, Deshpande JV, Lau EHY et al (2008) A test for constant fatality rate of an emerging epidemic: with applications to severe acute respiratory syndrome in Hong Kong and Beijing. Biometrics 64:869–876. doi:10.1111/j.1541-0420.2007.00935.x

Leung GM, Hedley AJ, Ho LM et al (2004) The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients. Ann Intern Med 141:662–673

Lunn M, McNeil D (1995) Applying Cox regression to competing risks. Biometrics 51:524–532. doi:10.2307/2532940

Murphy SA, Van der Vaart AW (2000) On profile likelihood (with discussion). J Am Stat Assoc 95:449–485. doi:10.2307/2669386

Rothman KJ (2002) Epidemiology: an introduction. Oxford University Press, New York

Tsodikov A, Garibotti G (2007) Profile information matrix for nonlinear transformation models. Lifetime Data Anal 13:139–159. doi:10.1007/s10985-006-9023-z

WHO (2003a) Cumulative number of reported probable cases of Severe Acute Respiratory Syndrome (SARS) (Up to 11 July 2003). World Health Organization, Geneva

WHO (2003b) SARS case fatality ratio, incubation period. Severe Acute Respiratory Syndrome (SARS) Disease Outbreak Reported. Update 49. World Health Organization, Geneva

Yip PS, Lam KF, Lau EH et al (2005) A comparison study of real-time fatality rates: severe acute respiratory syndrome in Hong Kong, Singapore, Taiwan, Toronto and Beijing, China. J R Stat Soc A 168:233–243

Yu PL, Chan JSK, Fung WK (2006) Statistical exploration from SARS. Am Stat 60:81–91. doi:10.1198/000313006X91421