

# Semiparametric Estimation of Time-Varying Intervention Effects Using Recurrent Event Data

Jiajun Xu<sup>a</sup>, K. F. Lam<sup>\*a</sup>, Feng Chen<sup>b</sup>, Paul Milligan<sup>c</sup>, and Yin-Bun Cheung<sup>d,e</sup>

<sup>a</sup>*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong*

<sup>b</sup>*School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052*

<sup>c</sup>*Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK*

<sup>d</sup>*Center for Quantitative Medicine, Office of Clinical Science, Duke-NUS Medical School, Singapore*

<sup>e</sup>*Tampere Center for Child Health Research, University of Tampere and Tampere University Hospital, Finland*

April 6, 2017

## Abstract

We consider the estimation of the optimal interval between doses for interventions such as malaria chemoprevention and vaccine booster doses that are applied intermittently in infectious disease control. A flexible exponential-like function to model the time-varying intervention effect in the framework of Andersen-Gill model for recurrent event time data, is considered. The partial likelihood estimation approach is adopted and a large scale simulation study is carried out to evaluate the performance of the proposed method. A simple guideline for the choice of the optimal interval

---

\*Corresponding author. K.F. Lam, The University of Hong Kong, Pokfulam Road, Hong Kong.

*E-mail addresses:* hrntlkf@hku.hk

between successive doses is proposed. The methodology is illustrated with the analysis of data from a malaria chemoprevention trial.

*Key words:* Andersen-Gill model, Ghana, malaria, proportional intensity, recurrent events, vaccine, waning efficacy.

## 1 Introduction

The efficacy of preventive interventions against infectious disease often wanes over time, the intervention therefore needs to be re-administered to sustain protection (for example, pertussis vaccines [11], and vaccines [22], insecticide-treated bednets [20] and chemoprevention [21] in malaria control). Reliable surrogates of protection may not exist, for these interventions the optimal interval for re-dosing needs to be determined from data on the incidence of disease events. In this paper we describe an approach for estimating time varying efficacy for interventions that are applied intermittently, and for estimating the optimal interval between successive doses. We illustrate the use of the method with data from a malaria chemoprevention trial. Chemoprevention for malaria involves repeated administration of antimalarial drugs at intervals and is used to protect children and pregnant women in endemic areas. The choice of the optimal interval is an important public health question.

In most statistical analyses of interventions in infectious diseases, the intervention effect or protective efficacy is generally assumed to be constant. Randomized trials of interventions against infectious disease are also often analyzed using data on first episode of the disease even when subsequent episodes have been recorded [1], leading to information loss and over-emphasis of short-term effects [10, 6]. The estimation of constant intervention effects using recurrent event data has been considered [6, 23, 24]. Previous studies have considered the role of frailty in the analysis of first or single disease episode in generating an apparent declining efficacy [9]. An advantage of analysis of recurrent events that has not been sufficiently discussed in this context is that, since a person remains in the analysis risk set after experiencing one or more events, frailty has little impact on generating an apparent declining efficacy [6, 24]. Methods for analysis of time-varying effects have been described using cubic splines [19, 12] and smoothed residuals [7]. These methods do not specifically address applications where interventions are repeatedly administered. They do not capture the sharp changes in hazard ratio at the times of booster or repeated doses. In this paper, our objectives are to model the time-varying efficacy of preventive intervention using recurrent event data and to determine the optimal interval between successive intervention doses so as to sustain a high level of protection.

Two issues need to be taken into consideration to establish an appropriate

statistical model for the analysis. Firstly, when a treatment is administered, there may be a residual effect from previous treatments that needs to be taken into account. Secondly, it is well known that individuals exposed to infection will acquire some level of immunity to the disease. Hence, in the context of infectious disease control, it may be unavoidable that interventions successful in preventing infection also prevent natural acquisition of immunity. When the intervention effect wears off, treated individuals may be more susceptible to infection than they would have been if they had not received the intervention. Hence, it is important to determine whether the preventive treatment will have a negative effect in the long run. Other preventive interventions that are administered intermittently for the control of infectious diseases would probably have similar concerns [15].

The remainder of this paper is organized in the following. The typical type of data under consideration, the proposed model, the estimation method, and a criterion for the determination of the optimal interval between successive preventive treatments will be discussed in Section 2. In Section 3, a large scale simulation study is conducted to assess the performance of the estimation method. In Section 4, the proposed model and method are applied to the data from a randomized chemoprevention trial for malaria in young children. Section 5 concludes with a discussion on the potential practical applications of this method.

## 2 The data, model and estimation method

Suppose we have a sample of  $n$  individuals who are randomized to either the intervention or control group. Define  $z_i$  as the group indicator with  $z_i = 1$  if subject  $i$  is in the intervention group and  $z_i = 0$  otherwise. Subjects allocated to the control group will receive no preventive intervention. Let  $0 < t_{i1} < t_{i2} < \dots$  denote the multiple onset times of the disease episodes experienced by the  $i^{\text{th}}$  subject, and  $0 \leq d_{i1} < d_{i2} < \dots < d_{ik}$  be the  $k$  pre-determined times for dosing, with treatment or placebo, to the  $i^{\text{th}}$  subject for  $i = 1, 2, \dots, n$ . Let  $\mathbf{x}_i(\cdot)$  denote the covariate vector process, and  $\tau_i$  the total follow-up time (censoring time) associated with the  $i^{\text{th}}$  subject. We assume independent censoring, that is, the censoring time  $\tau_i$  is independent of the onset times  $t_{ij}$ 's given the history of the covariate vector process  $\mathbf{x}_i(\cdot)$ . The covariate process is assumed to be left continuous with right-hand limits. In practice, it is often the case that a time-varying component of the covariate process is observed only at study entry and exit and at the event times. For practical purposes, we simply assume these covariates to be linear or constant between successive event times. Let  $N_i(t) = \sum_{j=1}^{\infty} \mathbf{1}\{t_{ij} \leq t\}$  denote the episode occurrence counting process of subject  $i$  and  $\mathcal{F}_t = \sigma\{z_i, \mathbf{x}_i(s), N_i(s), s \leq t, d_{ij}, j = 1, 2, \dots : d_{ij} \leq t; i = 1, \dots, n\}$  denote the  $\sigma$ -algebra generated by all the data by time  $t$ . It has been suggested that the

Andersen-Gill model [3] is useful for assessing the impact of preventive medicine on recurrent events in effectiveness studies [6]. As a measure of the individual risk of a new event, the intensity process of  $N_i(t)$  is assumed to have the form

$$\lambda_i(t | \mathcal{F}_{t-}) = \lambda_0(t) \exp[\boldsymbol{\beta}^\top \mathbf{x}_i(t) - z_i \cdot G(t)], \quad (1)$$

where  $\lambda_0(\cdot) \geq 0$  is an unspecified baseline intensity function,  $\boldsymbol{\beta}$  is the vector of regression coefficients and  $G(\cdot)$  is the function of the time-varying effect of the intervention.

One approach would be using a single function  $G(t)$  to represent the logarithm of the time-varying relative risk during the whole follow-up period, where  $t$  is the time since the first dose. However, in practice, it is difficult to arrange for each participant to receive the intervention strictly according to the designed schedule, and the variation in time to receive each dose among the subjects may lead to a biased evaluation of the intervention effect if this modeling approach is adopted. To account for the variations in the  $d_{ij}$ 's among the subjects, we assume

$$G(t) = \sum_{j:d_{ij}<t} g(t - d_{ij}) \quad (2)$$

where  $g(t - d_{ij})$  indicates the amount of reduction in the disease risk on the logarithm scale due to the  $j^{\text{th}}$  dose of intervention,  $(t - d_{ij})$  units of time after the administration of the  $j^{\text{th}}$  dose. Note that the effects due to the  $j^{\text{th}}$  and  $l^{\text{th}}$  doses ( $j \neq l$ ) are assumed to be additive on the log-scale. Then  $G(t)$  is the cumulative intervention effect for subject  $i$  at time  $t$ . For example, let  $d_{i2} < t \leq d_{i3}$ , then the cumulative intervention effect for subject  $i$  at time  $t$  is

$$G(t) = g(t - d_{i1}) + g(t - d_{i2}),$$

which is interpreted as the intervention effect of the second dose  $(t - d_{i2})$  units of time after the administration of it plus the residual intervention effect of the first dose  $(t - d_{i1})$  units of time after its administration. A graphical illustration of this example is given in Figure 1 and that optimization of the time between  $d_2$  and  $d_1$  is an important part of the analysis.

The exponential decay function,  $g(t) = a \exp(-bt)$ , has been used to model waning vaccine efficacy [11]. To answer the aforementioned research questions, we modify the exponential decay function as

$$g(t) = a \exp(-bt^\delta) + c, \quad \text{for } a, b, \delta \geq 0 \text{ and } -\infty < c < \infty. \quad (3)$$

The parameter  $a$  represents the dose effect immediately after each intervention while  $b$  controls the overall decline rate of  $g(t)$ . A negative value of  $c$  suggests that

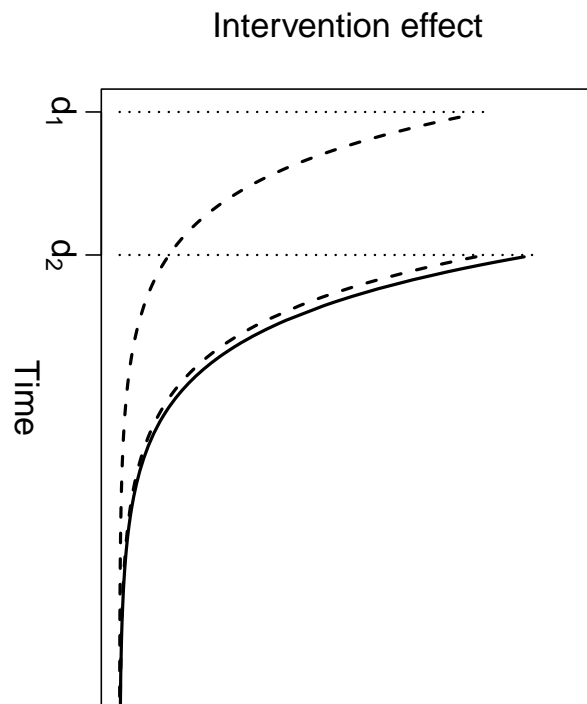


Figure 1: A graphical illustration of the cumulative intervention effect ( $G$ ). **solid line**: cumulative intervention effect for  $t > d_2$ ; **dashed lines**: intervention effects ( $g$ ) of the first and second doses.

the long-term intervention effect is harmful, which accommodates the possibility that the intervention hinders acquisition of natural immunity. The exponential decay function is a special case of our proposed function when  $\delta = 1$  and  $c = 0$ . The model we propose includes both initially rapid ( $\delta \leq 1$ ) and initially slow ( $\delta > 1$ ) decay as special cases. In pharmacodynamics, the relationship between physiological effect and drug concentration is often non-linear with an upper asymptote [8]. Given a sufficient dose such that the drug concentration immediately after each dose is near the asymptote level, one would expect  $\delta > 1$ .

The estimation of the model parameters  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$  where  $\boldsymbol{\gamma} = (a, b, c, \delta)^\top$  can be made possible based on the partial likelihood

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \prod_{j=1}^{N_i(\tau_i)} \frac{\exp\left(\boldsymbol{\beta}^\top \mathbf{x}_i(t_{ij}) - z_i \sum_{l:d_{il} < t_{ij}} g(t_{ij} - d_{il}; \boldsymbol{\gamma})\right)}{\sum_{k=1}^n Y_k(t_{ij}) \exp\left(\boldsymbol{\beta}^\top \mathbf{x}_k(t_{ij}) - z_k \sum_{l:d_{kl} < t_{ij}} g(t_{ij} - d_{kl}; \boldsymbol{\gamma})\right)} \quad (4)$$

where  $Y_i(t) = \mathbf{1}\{\tau_i \geq t\}$  is the at-risk indicator process. On letting  $\partial_{\boldsymbol{\gamma}} = \partial/\partial \boldsymbol{\gamma}$  and  $\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}^\top}^2 = \partial^2/(\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top)$ , the corresponding log likelihood function, score function, and the observed information matrix are respectively,

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^{N_i(\tau_i)} \left\{ \boldsymbol{\beta}^\top \mathbf{x}_i(t_{ij}) - z_i \cdot G_i(t_{ij}; \boldsymbol{\gamma}) - \log(S^{(0)}(t_{ij}; \boldsymbol{\theta})) \right\}, \quad (5)$$

$$U(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^{N_i(\tau_i)} \left\{ \begin{pmatrix} \mathbf{x}_i(t_{ij}) \\ -\partial_{\boldsymbol{\gamma}} z_i G_i(t_{ij}; \boldsymbol{\gamma}) \end{pmatrix} - \frac{S^{(1)}(t_{ij}; \boldsymbol{\theta})}{S^{(0)}(t_{ij}; \boldsymbol{\theta})} \right\}, \quad (6)$$

$$I(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^{N_i(\tau_i)} \left\{ \begin{pmatrix} 0 & 0 \\ 0 & \partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}^\top}^2 z_i G_i(t_{ij}; \boldsymbol{\gamma}) \end{pmatrix} + \frac{S^{(2)}(t_{ij}; \boldsymbol{\theta})}{S^{(0)}(t_{ij}; \boldsymbol{\theta})} - \frac{S^{(1)}(t_{ij}; \boldsymbol{\theta})^{\otimes 2}}{S^{(0)}(t_{ij}; \boldsymbol{\theta})^2} \right\} \quad (7)$$

where

$$G_i(t; \boldsymbol{\gamma}) = \sum_{l:d_{il} < t} g(t - d_{il}; \boldsymbol{\gamma}), \quad (8)$$

$$S^{(0)}(t; \boldsymbol{\theta}) = \sum_{k=1}^n Y_k(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}_k(t) - z_k \cdot G_k(t; \boldsymbol{\gamma})), \quad (9)$$

$$S^{(1)}(t; \boldsymbol{\theta}) = \sum_{k=1}^n Y_k(t) \begin{pmatrix} \mathbf{x}_k(t) \\ -\partial_{\boldsymbol{\gamma}} z_k G_k(t; \boldsymbol{\gamma}) \end{pmatrix} \exp(\boldsymbol{\beta}^\top \mathbf{x}_k(t) - z_k \cdot G_k(t; \boldsymbol{\gamma})), \quad (10)$$

$$S^{(2)}(t; \boldsymbol{\theta}) = \sum_{k=1}^n Y_k(t) \left[ \begin{pmatrix} \mathbf{x}_k(t) \\ -\partial_{\boldsymbol{\gamma}} z_k G_k(t; \boldsymbol{\gamma}) \end{pmatrix}^{\otimes 2} + \begin{pmatrix} 0 & 0 \\ 0 & -\partial_{\boldsymbol{\gamma}\boldsymbol{\gamma}^\top}^2 z_k G_k(t; \boldsymbol{\gamma}) \end{pmatrix} \right] \exp(\boldsymbol{\beta}^\top \mathbf{x}_k(t) - z_k \cdot G_k(t; \boldsymbol{\gamma})) \quad (11)$$

The expressions for the derivatives of  $g$  in (6) - (7) depend on the form of  $g$ . In the case of the exponential-like function as in (3), the derivatives are as follows,

$$\frac{\partial}{\partial \boldsymbol{\gamma}} g(t; \boldsymbol{\gamma}) = (\exp(-bt^\delta), -at^\delta \exp(-bt^\delta), 1, -abt^\delta \log(t) \exp(-bt^\delta))^\top,$$

$$\frac{\partial^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} g(t; \boldsymbol{\gamma}) = \begin{pmatrix} 0 & -t^\delta \exp(-bt^\delta) & 0 & -bt^\delta \log(t) \exp(-bt^\delta) \\ -t^\delta \exp(-bt^\delta) & at^{2\delta} \exp(-bt^\delta) & 0 & -at^\delta \log(t) \exp(-bt^\delta) (1 - bt^\delta) \\ 0 & 0 & 0 & 0 \\ -bt^\delta \log(t) \exp(-bt^\delta) & -at^\delta \log(t) \exp(-bt^\delta) (1 - bt^\delta) & 0 & -abt^\delta (\log(t))^2 \exp(-bt^\delta) (1 - bt^\delta) \end{pmatrix}.$$

The maximum partial likelihood estimator (MPLE)  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  can be obtained by maximizing the log likelihood function  $\ell(\boldsymbol{\theta})$ , or equivalently solving the score equation  $U(\boldsymbol{\theta}) = 0$  using, for example, the Newton-Raphson algorithm, whereby the following is iterated until convergence,

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + I(\boldsymbol{\theta})^{-1} U(\boldsymbol{\theta}). \quad (12)$$

The cumulative baseline intensity function  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$  can be estimated by the Breslow estimator [4] given by

$$\hat{\Lambda}_0(t) = \int_0^t \frac{dN(u)}{S^{(0)}(u; \hat{\boldsymbol{\theta}})}, \quad (13)$$

where  $N(t) = \sum_{i=1}^n N_i(t \wedge \tau_i)$ .

Under regularity conditions similar to VII.2.1 and VII.2.2 in [2], it can be shown that, as the sample size  $n$  tends to infinity, the MPLE  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  is consistent and asymptotically normal with its asymptotic variance consistently estimated by the inverse of the observed information matrix evaluated at the MPLE, and the Breslow estimator of  $\Lambda_0(\cdot)$  is consistent and that  $\sqrt{n}\{\hat{\Lambda}_0(\cdot) - \Lambda_0(\cdot)\}$  converges to a Gaussian process with zero mean and independent increments.

Under the framework of the Andersen-Gill model, the protective efficacy at time  $t$  due to the intervention is defined as

$$\text{PE}(t) = 1 - \frac{\lambda(t \mid z = 1, \boldsymbol{x}(t))}{\lambda(t \mid z = 0, \boldsymbol{x}(t))} = 1 - \exp[-G(t)],$$

which can be interpreted as the percentage reduction in the risk of disease at time  $t$  due to the intervention and we denote  $\text{PE}(d_1) = \text{PE}_0$  be the protective efficacy immediately after the first dose. Because of its simple nature, we propose to determine a target interval between successive supplementary interventions such that the protective efficacy of the intervention will be retained at a high level relative to  $\text{PE}_0$  at all time, say  $\text{PE}(t) \geq \zeta \cdot \text{PE}_0$  for  $d_1 < t \leq d_k$  where  $0 < \zeta < 1$  is a pre-specified fraction. More discussion on the target interval is given in Section 4 with an application to the randomized chemoprevention trial for malaria.

### 3 Simulation study

We assess the performance of the MPLE  $\hat{\theta}$  through a simulation study. In the simulation, we mimic the malaria trial example to be considered in the next section by assuming the covariates vector  $\mathbf{x}_i(t) \equiv \mathbf{x}_i = (X_{i1}, X_{i2})^\top$  where  $X_{i1}$  are generated independently from a Bernoulli distribution with  $P(X_{i1} = 1) = P(X_{i1} = 0) = 0.5$ , and  $X_{i2}$  are from the uniform distribution over the interval  $(0, 1)$ , mimicking the gender and age at entry of the subjects, respectively. The regression coefficients are set to be  $\beta = (-0.5, 0.5)^\top$ . We let  $Z$  be the intervention indicator with  $Z = 1$  for the intervention group and  $Z = 0$  for the control group, respectively.  $Z_i$ 's are generated from a Bernoulli distribution with  $P(Z_i = 1) = 0.5$ . Each subject will receive 4 doses of intervention/placebo in total. Being consistent with the malaria data set, we let the clock start at the time when the first dose is given, so that  $d_{i1} \equiv 0$  for  $i = 1, \dots, n$ . To accommodate the variation in time to receive doses in the subsequent interventions, we let the times to receive the second to the fourth dose for subject  $i$  be generated independently from  $d_{i2} \sim U(0.8, 1.2)$ ,  $d_{i3} \sim U(5.8, 6.2)$ ,  $d_{i4} \sim U(8.8, 9.2)$ , respectively. The censoring times are set to be  $\tau_i = 24$ . The intervention effect function as in (3) with  $a = 2$  and  $b = 0.5$  is considered throughout the simulation study. Three different values for  $c$ , namely  $-0.04$ ,  $0$  and  $0.04$ , are considered to mimic a negative, neutral and positive long-term intervention effect, respectively. Two values, namely  $\delta = 1$  and  $\delta = 3$ , are considered to mimic different declining patterns of the intervention effect. To simplify the data generation, the baseline intensity is set to be  $\lambda_0(t) = 0.06$  so that each subject will experience 1.3 episodes on the average in the simulated data set. The sample size  $n$  is set to be  $n = 1000, 1500$  or  $2000$ .

The generation of recurrent event times usually involves finding the cumulative intensity function  $\Lambda_i(t)$ . In the presence of time dependent components in the model, the calculation requires numerical integration. To avoid this, we considered the thinning approach [13] to generate the recurrent event times. Briefly, we can choose a value  $\bar{\lambda}$  such that for  $\forall t$ , the intensity function satisfies  $\lambda(t) \leq \bar{\lambda}$ . From our experience, the choice of  $\bar{\lambda} = 1.05 \left[ \max_t \lambda(t) \right]$  is quite efficient in generating the recurrent event times. If a set of times  $T_1^*, T_2^*, \dots$  are random samples generated from a homogeneous Poisson process with constant rate  $\bar{\lambda}$ , then each time  $T_j^*$  has a probability of  $\lambda(T_j^*)/\bar{\lambda}$  to be an actual event time from the process with intensity  $\lambda(t)$ . In this study, for each subject  $i$  with their particular intensity function  $\lambda_i(t)$ , we can generate the event times according to the following steps:

1. Set  $T_{i0} = 0, T^* = 0$  and  $j = 1$ .
2. Draw a random number  $R \sim Exp(\bar{\lambda})$ , and update  $T^* = T^* + R$ .
3. Generate  $V \sim U(0, 1)$ :



- If  $V \leq \frac{\lambda_i(T^*)}{\bar{\lambda}}$ ,  $T^*$  is an event time, then we set  $T_{ij} = T^*$  as the  $j^{th}$  event time of subject  $i$ , update  $j = j + 1$  and return to step 2.
- If  $V > \frac{\lambda_i(T^*)}{\bar{\lambda}}$ ,  $T^*$  is not an event time, we directly go back to step 2 without updating  $j$  and  $T_{ij}$ .

For each combination of the sample size  $n$  and the intervention effect function  $g(\cdot)$ , the covariates and dosing times for all the  $n$  subjects are generated once and held fixed, the event times for each subject are then simulated according to the intensity model. The parameters  $\theta$  are then estimated using the MPLE. The data simulation and parameter estimation process is repeated for a total of 500 times. For each combination of the values of  $n$ ,  $c$  and  $\delta$ , the average of the 500 estimates (mean.est), the average of the 500 estimated standard errors (se), the empirical coverage of the 95% confidence interval based on the asymptotic properties of the estimator (cp) and the empirical standard deviation of the 500 estimates (esd) are computed for each parameter. The results are summarized in Tables 1 to 2.

The performance of the estimator is shown to be highly satisfactory from the simulation studies. The mean estimates for the parameters are generally quite close to their respective true values, and the average estimated standard errors closely resembled the empirical standard deviations of the estimates. When the sample size is small, the mean estimates for  $a$  and  $b$  seem to be slightly larger than the true values, but the difference is not alarming with respect to the large standard error estimates. Nevertheless, this phenomenon diminishes as  $n$  gets larger. The finite sample properties of the estimator are well justified as the empirical coverage probabilities are all reasonably close to the nominal level 0.95.

Figure 2 and Figure 3 show the point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the 500 estimates for  $g(t)$  together with the true  $g(t)$ . Generally, the estimates are able to reveal the true shape of the intervention effects function, as we can see that the  $50^{th}$  percentile of the estimated  $g(t)$  coincides with the true  $g(t)$ . The true function of  $g(t)$  presented in Figure 2 is the exponential decay function that is used frequently in clinical studies for describing declining intervention effects. Our proposed function is a generalized form of this exponential decay function and the proposed method is able to estimate it quite well.

To assess the performance of the estimator  $\hat{\Lambda}_0(t)$  for the cumulative baseline intensity function, the point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the 500 estimates for  $\Lambda_0(t)$  together with the true baseline intensity are presented in Figures 4 to 5. It is reasonable to claim that the performance of  $\hat{\Lambda}_0(t)$  is highly satisfactory as the cumulative baseline intensity is accurately estimated although we also note that the variation of the estimates grows with  $t$ , but this is just a natural property of the Nelson-Aalen or Breslow estimator [4, 2].

Table 1: Simulation results based on the proposed model with  $\delta = 1$

$\delta = 1$														
	true	mean.est	$n = 1000$			$n = 1500$				$n = 2000$				
			se	esd	cp	mean.est	se	esd	cp	mean.est	se	esd	cp	
$c = -0.04$	$\beta_1$	-0.500	-0.501	0.059	0.057	0.961	-0.498	0.048	0.049	0.939	-0.499	0.042	0.042	0.962
	$\beta_2$	0.500	0.496	0.099	0.097	0.957	0.499	0.081	0.080	0.953	0.499	0.070	0.067	0.968
	$a$	2.000	2.379	1.149	1.189	0.913	2.202	0.716	0.729	0.921	2.146	0.560	0.590	0.935
	$b$	0.500	0.582	0.394	0.395	0.886	0.549	0.303	0.313	0.890	0.536	0.255	0.261	0.911
	$c$	-0.040	-0.053	0.040	0.041	0.966	-0.047	0.028	0.028	0.972	-0.043	0.022	0.021	0.962
	$\delta$	1.000	1.116	0.507	0.549	0.943	1.079	0.392	0.408	0.955	1.057	0.329	0.347	0.943
$c = 0$	$\beta_1$	-0.500	-0.502	0.061	0.065	0.943	-0.504	0.049	0.048	0.951	-0.499	0.043	0.044	0.931
	$\beta_2$	0.500	0.502	0.101	0.103	0.947	0.494	0.084	0.084	0.945	0.501	0.073	0.072	0.960
	$a$	2.000	2.505	1.362	1.476	0.955	2.290	0.926	1.063	0.943	2.120	0.568	0.553	0.952
	$b$	0.500	0.608	0.421	0.433	0.914	0.561	0.330	0.362	0.916	0.529	0.264	0.256	0.939
	$c$	0.000	-0.012	0.040	0.039	0.971	-0.008	0.030	0.029	0.980	-0.003	0.023	0.023	0.968
	$\delta$	1.000	1.091	0.502	0.517	0.939	1.079	0.406	0.409	0.939	1.061	0.339	0.326	0.970
$c = 0.04$	$\beta_1$	-0.500	-0.505	0.062	0.065	0.951	-0.502	0.050	0.049	0.957	-0.502	0.044	0.042	0.965
	$\beta_2$	0.500	0.503	0.106	0.099	0.969	0.495	0.085	0.083	0.949	0.502	0.075	0.073	0.953
	$a$	2.000	2.425	1.362	1.452	0.906	2.251	0.850	0.868	0.945	2.220	0.682	0.676	0.955
	$b$	0.500	0.563	0.423	0.440	0.857	0.562	0.335	0.330	0.937	0.568	0.289	0.294	0.941
	$c$	0.040	0.031	0.039	0.039	0.957	0.033	0.031	0.030	0.970	0.033	0.026	0.027	0.965
	$\delta$	1.000	1.226	0.612	0.720	0.947	1.063	0.410	0.392	0.957	1.031	0.346	0.358	0.933

Table 2: Simulation results based on the proposed model with  $\delta = 3$

$\delta = 3$														
	true	mean.est	$n = 1000$			$n = 1500$				$n = 2000$				
			se	esd	cp	mean.est	se	esd	cp	mean.est	se	esd	cp	
$c = -0.04$	$\beta_1$	-0.500	-0.501	0.056	0.057	0.952	-0.501	0.045	0.047	0.936	-0.501	0.039	0.038	0.951
	$\beta_2$	0.500	0.506	0.093	0.092	0.943	0.499	0.074	0.072	0.955	0.500	0.066	0.066	0.951
	$a$	2.000	2.185	0.506	0.570	0.931	2.072	0.348	0.366	0.961	2.055	0.290	0.303	0.957
	$b$	0.500	0.561	0.321	0.387	0.853	0.512	0.260	0.290	0.881	0.512	0.225	0.235	0.918
	$c$	-0.040	-0.041	0.017	0.017	0.941	-0.041	0.014	0.013	0.949	-0.041	0.012	0.012	0.947
	$\delta$	3.000	4.092	2.392	3.141	0.845	3.747	1.764	2.084	0.910	3.417	1.344	1.388	0.929
$c = 0$	$\beta_1$	-0.500	-0.500	0.057	0.057	0.938	-0.502	0.047	0.045	0.949	-0.500	0.040	0.041	0.945
	$\beta_2$	0.500	0.496	0.098	0.096	0.961	0.497	0.081	0.080	0.947	0.501	0.068	0.066	0.949
	$a$	2.000	2.136	0.498	0.565	0.942	2.076	0.363	0.363	0.945	2.072	0.315	0.324	0.955
	$b$	0.500	0.524	0.324	0.365	0.847	0.512	0.267	0.283	0.896	0.535	0.242	0.248	0.928
	$c$	0.000	-0.002	0.017	0.017	0.957	0.000	0.014	0.015	0.933	-0.001	0.012	0.012	0.947
	$\delta$	3.000	4.305	2.647	3.595	0.884	3.855	1.884	2.389	0.912	3.328	1.390	1.420	0.914
$c = 0.04$	$\beta_1$	-0.500	-0.501	0.059	0.060	0.954	-0.501	0.048	0.048	0.957	-0.501	0.042	0.043	0.947
	$\beta_2$	0.500	0.505	0.101	0.095	0.958	0.498	0.082	0.084	0.942	0.491	0.071	0.068	0.953
	$a$	2.000	2.204	0.593	0.736	0.947	2.098	0.397	0.422	0.957	2.052	0.323	0.310	0.959
	$b$	0.500	0.563	0.356	0.399	0.876	0.532	0.291	0.310	0.896	0.505	0.249	0.256	0.901
	$c$	0.040	0.038	0.018	0.018	0.945	0.040	0.015	0.015	0.946	0.040	0.013	0.013	0.947
	$\delta$	3.000	3.996	2.383	3.060	0.880	3.674	1.926	2.039	0.923	3.561	1.618	1.780	0.932

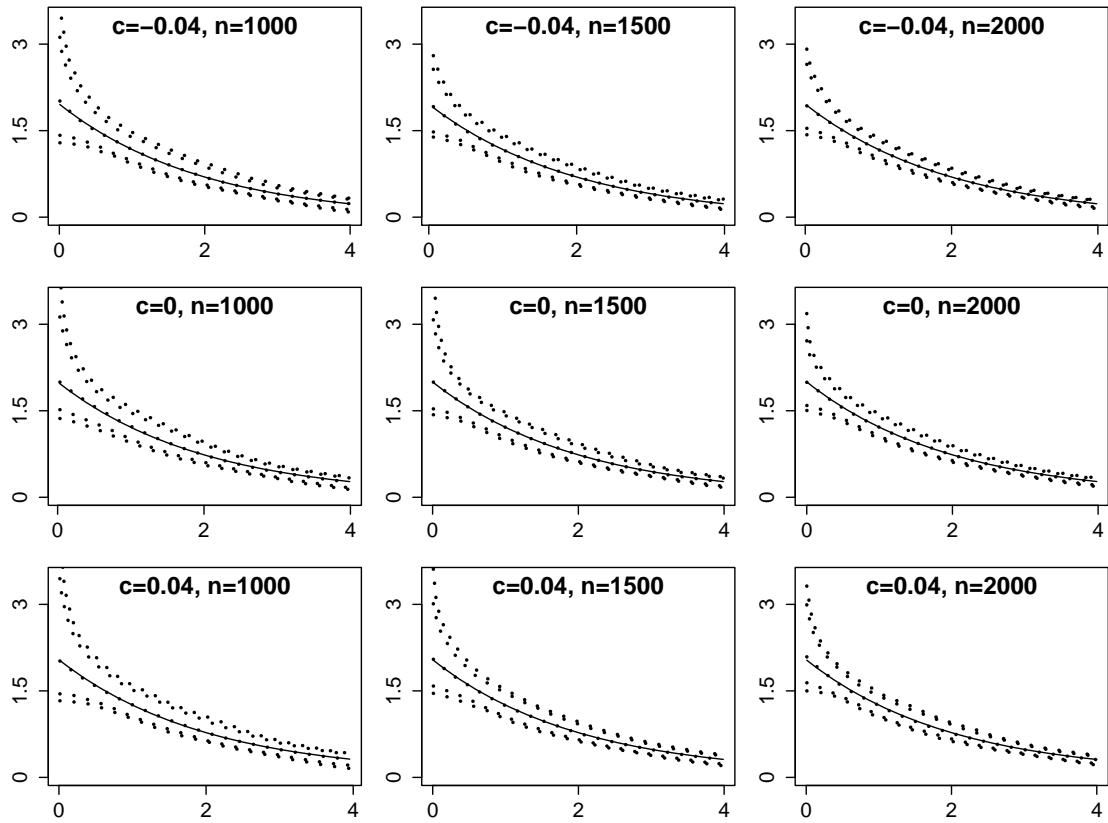


Figure 2: Plot of the true  $g(t)$  and the empirical point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the estimates for  $g(t)$  when  $\delta = 1$ .

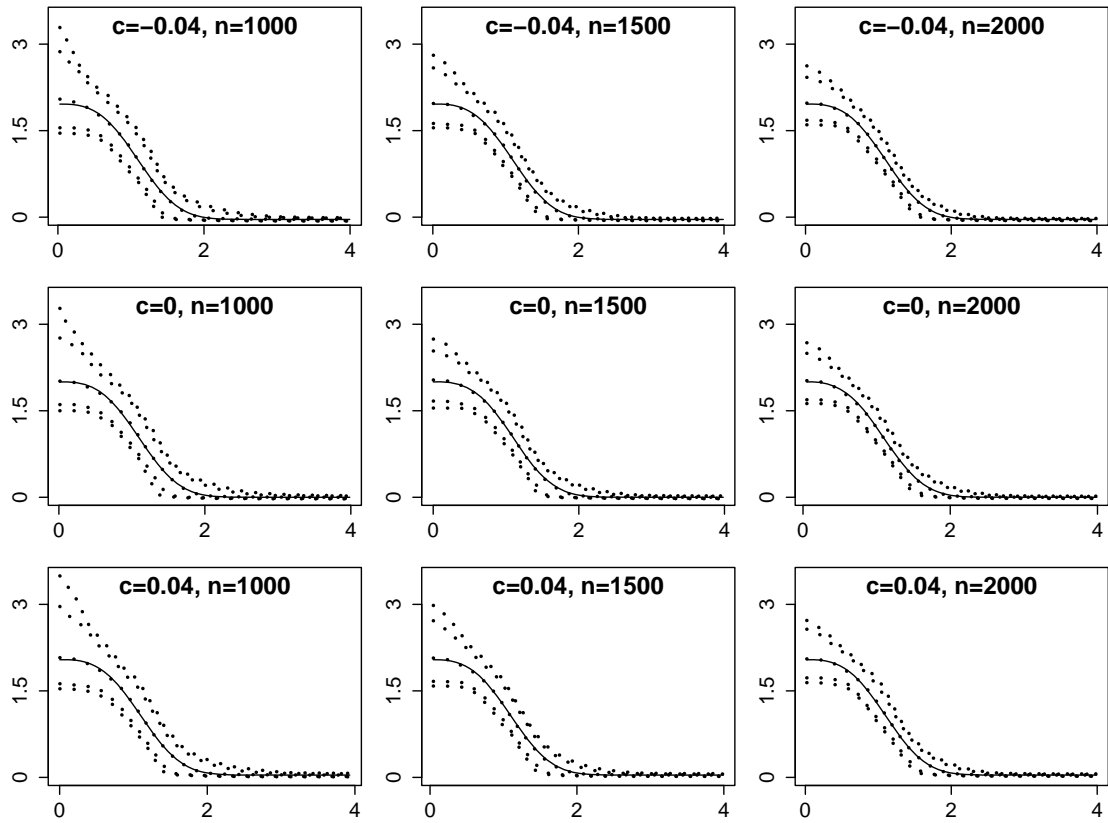


Figure 3: Plot of the true  $g(t)$  and the empirical point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the estimates for  $g(t)$  when  $\delta = 3$ .

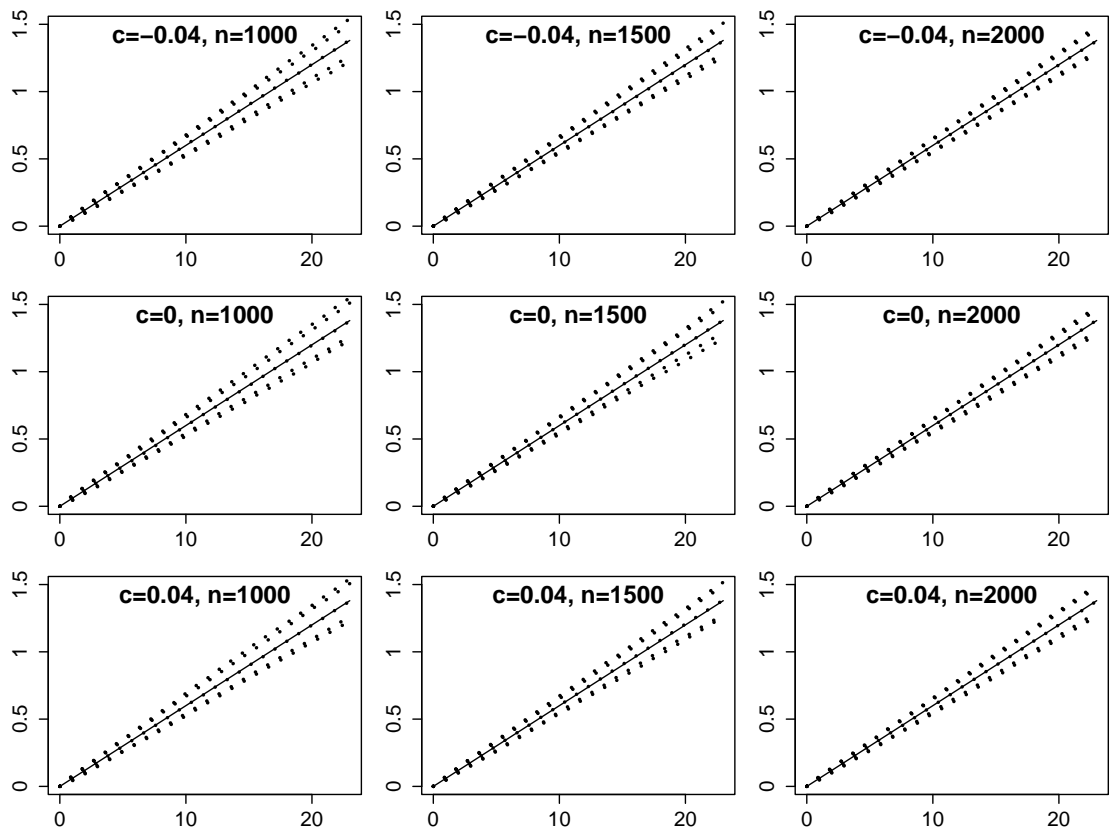


Figure 4: Plot of the true  $\Lambda_0(t)$  and the empirical point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the estimates for  $\Lambda_0(t)$  when  $\delta = 1$ .

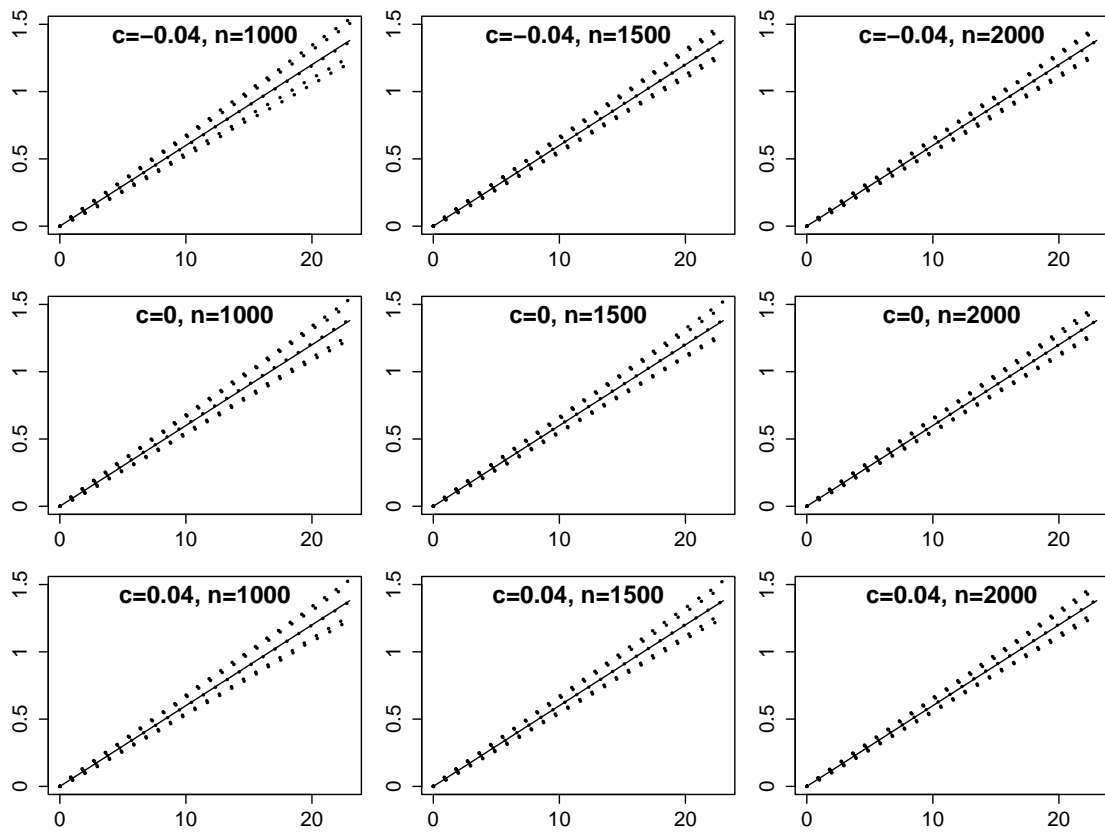


Figure 5: Plot of the true  $\Lambda_0(t)$  and the empirical point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the estimates for  $\Lambda_0(t)$  when  $\delta = 3$ .

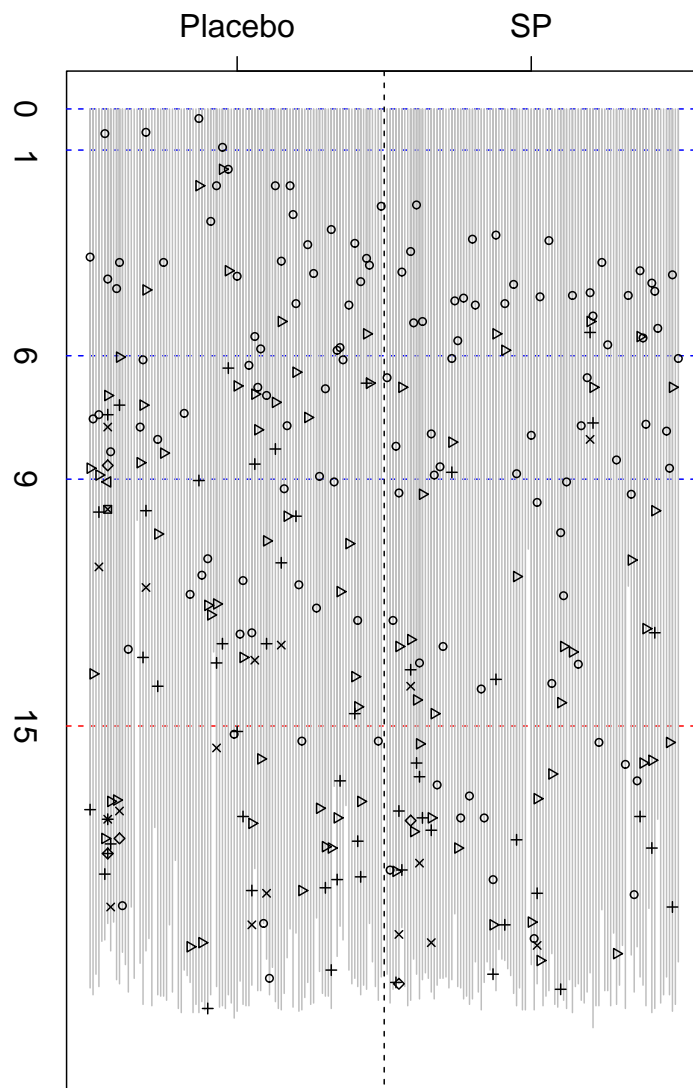
## 4 Application to the malaria data

We consider data from a randomized placebo-controlled trial of chemoprevention for malaria in young children conducted in Ghana between September, 2000 and June, 2004 [5]. A total of 2485 infants with mean age 2 months (s.d. = 0.61 month) were enrolled when they attended the immunization clinic for a dose of diphtheria-pertussis-tetanus (DPT) vaccine. They were randomized to receive placebo or sulfadoxine-pyrimethamine (SP) to prevent malaria at their subsequent immunization visits. The participants were scheduled to receive 4 doses of intervention/placebo intermittently, namely, at 1 month, 2 months, 7 months and 10 months after enrolment. Since the intervention effect of SP is the main issue to be investigated in the study, we choose the date at which the first dose of SP/placebo was given to each subject as the origin of time at risk in the analysis. Tests for malaria were performed on children attending the health facilities with a febrile illness. Cases with malaria parasites in the blood confirmed by microscopy, together with history of fever or temperature  $\geq 37.5$  degree Celsius were defined as cases of malaria. To avoid double counting of the same malaria episode, cases recorded in children who returned to the clinic with malaria within 7 days after an initial diagnosis of malaria were ignored. The planned follow-up time for all children was  $\tau = 24$  months. Some participants were lost to follow up during the study period and were treated as censored observations at which the censoring times were the times at last follow-up. 89% (1103/1242) of the children in the placebo group and 88% (1088/1243) of the children in the SP group completed the full follow-up period. Among all those infants, 71.5% of them had at least one malaria episode during the follow-up.

The covariates considered in this study are infants' age at baseline (age) and the seasonal transmission of malaria (season, 1: rainy, 0: dry). Most cases of malaria occurred during the rainy season (July to November in Ghana), coinciding with the seasonal increase in the mosquito population. Since the participants were followed for a nearly two-year time span, season must be considered as a time dependent variable.

Figure 6 shows the event times of a randomly selected sample (100 individuals from both groups) from the data. Through a visual inspection of the density of event time dots between the two groups, there is some evidence that the number of events in the intervention group is much smaller than that in the control group in the month right after each intervention administered to subjects of the intervention group, but the numbers are similar about one month after each intervention, consistent with the rapid waning of the intervention effect. Moreover, if we focus on the intervention effect in the long run, say from 15 months after the first intervention, it seems that we can find more malaria episodes in the intervention group than in the control group. Therefore, a model which can take into account

Figure 6: A random sample of the participants with their records of malaria infections during the follow-up.



○ - first malaria infection, △, +, ×, ◇ are respectively first, second, third, fourth recurrent infections; **solid line**: indicating the re-organized times (in month) for the 4 doses, namely  $t = 0$  for the first dose,  $t = 1$  for the second dose,  $t = 6$  for the third dose and  $t = 9$  for the fourth dose; **dashed line**: indicating the time  $t = 15$  since the first dose.



Table 3: Estimation results for the malaria data based on the standard Andersen-Gill model and the proposed model.

	parameter	est	se
A-G	$Z$	-0.1380	0.0364
	season	1.5075	0.0437
	age	0.1090	0.0286
MPLE	$a$	1.5037	0.1591
	$b$	0.3959	0.1528
	$c$	-0.0220	0.0119
	$\delta$	3.8764	1.1939
	season	1.5115	0.0437
	age	0.1083	0.0285

long-term intervention effects may be necessary.

In the first analysis, we assumed the intervention effect to be constant by fitting the standard Andersen-Gill model to the data with  $Z = 1$  if the subject was in the intervention group and  $Z = 0$  otherwise. The results were summarized in the first part of Table 3. From this analysis, there is a 12.9 percent ( $\hat{\beta}_Z = -0.1380$ ; 95% CI=[-0.2093, -0.0667]) reduction in the risk of malaria among those who received intervention relative to those who received placebo, with adjustment to seasonal effect and age at enrollment. As expected, the rainy season is strongly associated with the risk of malaria ( $\hat{\beta}_{\text{season}} = 1.5075$ ; 95% CI=[1.4218, 1.5932]). We also observe that the risk of malaria is positively associated with infants' age at entry ( $\hat{\beta}_{\text{age}} = 0.1090$ ; 95% CI=[0.0529, 0.1651]).

To describe how the effect of SP varied over time and to assess the intervention effect in the long run, we apply our proposed model and the MPLE estimation method to the data. The estimates for the regression parameters are summarized in the second part of Table 3. It is seen that the estimates for wet season and age at entry are almost identical to those based on the standard Andersen-Gill model.

Figure 7 shows the plot of the logarithm of the hazard ratio (HR) of a treated individual against an untreated individual over time by keeping the season and age at entry at some fixed values. As expected, the graph indicates a steep reduction in the risk of malaria immediately after each dose, and the protective efficacy ( $1 - \text{HR}$ ) declined gradually over time. The estimated value of  $\delta$  was 3.88 that led to a relatively stable HR for about 1 month after each dose of SP. Interestingly, the logarithm of the HR immediately after the second dose was much lower than that after the first dose, consistent with a residual effect of the first intervention persisting until administering of the second intervention about a month later. There

was no evidence of a residual effect when the third dose was administered about 5 months later. From 2.5 to 3 months after each dose, the log HR curve started to go above the horizontal axis, indicating slightly increased risk of malaria infection among subjects in the intervention group. However, there was no clear evidence that the individuals became more vulnerable after receiving SP, as we can also see from Table 3 that the parameter  $c$  is not significantly different from zero at the 5% level.

The estimated cumulative baseline intensity functions for the two models are presented in Figure 8. Again, the two estimated functions are similar to each other, both showing a roughly linear growth rate in the baseline cumulative intensity over time.

Seasonality has a major impact on the risk of malaria disease. The risk of malaria in the dry season is only  $\exp(-1.5115) = 22\%$  of that in the rainy season. Notice that the estimated value of  $a$  is about the same as the estimated seasonal effect, meaning that during the rainy season, the chemoprevention is able to control the risk of the disease at the dry season level, at least in the short run. Hence these data support the policy of implementing chemoprevention only in rainy season. Suppose it is desirable to retain a high level of protective efficacy throughout the observation period such that  $PE(t) \geq \zeta \cdot PE_0$  for  $t > d_1$ , say for  $\zeta = 0.8$  or  $0.9$ , we can determine the optimal time for the second dose, namely  $d_2$ , such that

$$1 - \exp\{-g(d_2 - d_1; \gamma)\} = \zeta \cdot PE_0$$

is satisfied where  $PE_0 = 1 - \exp\{-G(d_1; \gamma)\}$  is the protective efficacy immediately after the first dose. Based on the estimated  $g(t; \gamma)$  and by setting  $\zeta = 0.8$ , it is easy to solve that  $d_2 = 30.5$  days after the first dose. Now given the second dose administered at time  $d_2 = 30.5$ , we can continue to find  $d_3$  based on the same rule, such that

$$1 - \exp\{-g(d_3 - d_1; \gamma) - g(d_3 - d_2; \gamma)\} = \zeta \cdot PE_0,$$

and we obtain  $d_3 = 60.7$  (30.2 days after the second dose) and  $d_4 = 90.5$  (29.8 days after the third dose) iteratively. We note that the obtained intervals,  $(d_j - d_{j-1})$  ( $j = 2, 3, 4$ ), are all very close to each other. Therefore, for a convenient implementation in real clinical practice, we may take 30 days as the *target* interval between 2 successive doses of SP in a sense that the protective efficacy of the intervention would be kept at least 80% of its initial level throughout. Moreover, by setting  $\zeta = 0.9$ , the optimal times are  $d_2 = 25.7$ ,  $d_3 = 52.1$  and  $d_4 = 77.9$  with the target interval around 26 days between any two successive doses of SP.

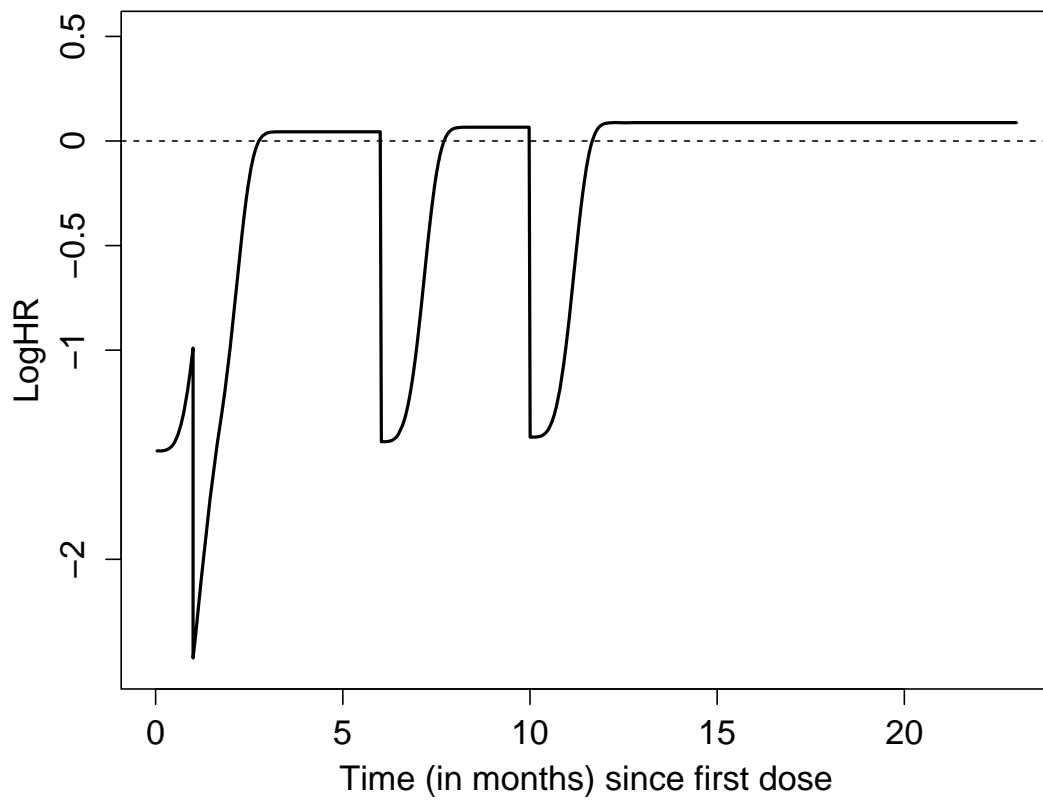


Figure 7: Plot of the estimated logarithm of the time-varying hazard ratio of a treated individual against an untreated individual based on the proposed method, with all other covariates being held fixed.

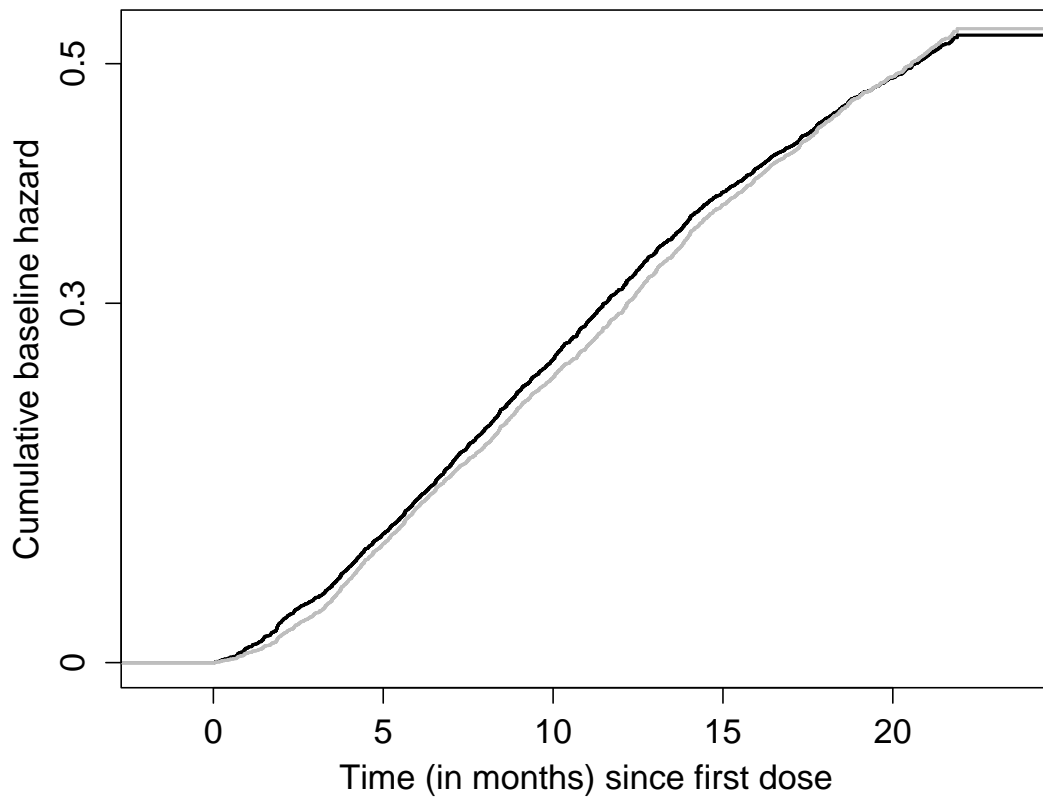


Figure 8: Plot of the estimated cumulative baseline hazard function based on the standard Andersen-Gill model and the proposed model. **black line**: estimated cumulative baseline hazard based on the proposed model; **grey line**: the estimated cumulative baseline hazard based on the standard Andersen-Gill model.

## 5 Discussion

We proposed a flexible exponential-like function to model the time-varying intervention effect of malaria chemoprevention in the framework of Andersen-Gill model. The incorporated parameter  $\delta$  allows for more flexible shapes in the modeled function, for example,  $\delta > 1$  indicates a slowly declining intervention effect during the early period after the intervention, while  $\delta \leq 1$  will lead to a rapidly declining effect shortly after the administration of the intervention. The cumulative intervention effect as described in (2) helps to explain the *enhanced protection* phenomenon observed in the malaria data set that malaria infections within one month after the second dose of SP were very rare. The parameter  $c$  is incorporated to account for the possible unintended long-term negative effect of the intervention due to reduced acquisition of natural immunity. As stated in [5] and [14], the intermittent interventions gave a high degree of protection against malaria infection, but may also impair natural acquisition of immunity to some degree, which might increase the risk after the intervention effect wanes. From our statistical findings, the estimated parameter  $c$  is a very small negative number that is not significantly different from 0. There might be a small negative long-term effect associated with repeated administration of preventive intervention, but this effect is outweighed by the benefits from the intervention. Furthermore, this negative effect will be expected to decay over time as those in the intervention group catch-up their natural immunity after the intervention has waned, a feature we did not include in the model because the study follow-up was only two years.

In another simulation study with similar setup as in Section 3, we let

$$g(t) = \begin{cases} -2t + 2 & 0 < t < 1.05, \\ -0.1 & t \geq 1.05 \end{cases}$$

to evaluate the flexibility and capability of the proposed model and method, even when the intervention is declining in a linear rate. The working model is based on  $g(t)$  as specified in (3). The true function, the mean estimated  $g(t)$  and some point-wise percentiles are plotted in Figure 9 which shows that the proposed method is able to approximate the shape of  $g(t)$  even if it is a straight line as the true function mostly falls in the center of the estimated curves.

We used a parametric function to model the potential time-varying intervention effect. The exponential function is widely used in medical applications to describe declining intervention efficacy, and it is convenient to define a threshold based on the parametric function as the time to give a booster dose. Therefore, in this study we consider the definition of the time-varying protective efficacy in the framework of Andersen-Gill model, that is,  $PE(t) = 1 - HR(t)$ , and we suggest that the time to give a booster dose of intervention is when the  $PE(t)$  starts to drop below a

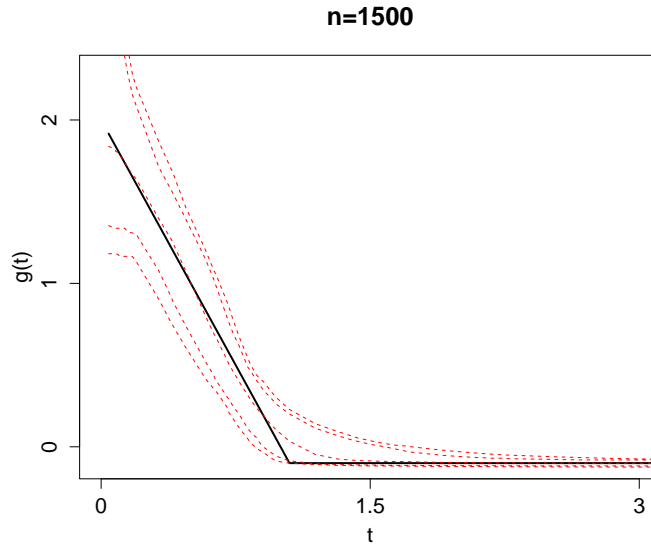


Figure 9: Plot of the approximation to the linear intervention function with sample size  $n = 1500$ . **solid line**: the true intervention function  $g(t)$ ; **red dashed line**: empirical point-wise percentiles ( $5^{th}$ ,  $10^{th}$ ,  $50^{th}$ ,  $90^{th}$  and  $95^{th}$ ) of the estimates for  $g(t)$ .

given fraction of its initial level, with this fraction being chosen by taking into account the feasibility and safety of more frequent intervention administrations.

Intermittent administration of antimalarial drugs to prevent malaria is increasingly used by malaria control programmes. Optimal protection requires that the interval between doses is carefully chosen to keep the individual at a low risk of disease. The methods we have described can be used to determine the target intervals between interventions needed to maintain a given minimum level of efficacy. The choice of interval in practice will be a balance between efficacy and the feasibility and tolerability of more frequent dosing. The proposed approach may also be applicable to other interventions that are administered intermittently, such as vaccines that require booster doses but for which good surrogate biomarkers of protection are not available.

Analysis of multiple recurrent events data is generally complicated by its dependence nature. In this paper, we specify the counting process  $N_i(t)$  of event recurrences for subject  $i$  as having an intensity process in Andersen-Gill's multiplicative form in (1). With a known dependence structure from the basic knowledge of the underlying problem, one may consider the time-shifted hazard form  $\lambda(t - S_{i,N_i(t-)})$  as discussed in [18] or one of the many other forms reviewed in [17]. Note that the treatment of the asymptotic theory for the estimator would also be different

if adopting a different model for  $\lambda$  such as the dynamical recurrent events model  $\lambda(t - S_{i,N_i(t-)})$  [16]. Nevertheless, the choice of the modelling approach depends on many other factors such as the background knowledge of the problem as well as the hypotheses to be addressed in practice.

### Acknowledgements

We acknowledge the valuable comments and suggestions of the Editor, an Associate Editor and two reviewers. We are grateful to Prof D Chandramohan for permission to use the data from the Ghana trial. The work of YBC was supported by the National Research Foundation, Singapore, under its Clinician Scientist Award (Award No. NMRC/CSA/0039/2012) administered by the Singapore Ministry of Health's National Medical Research Council.

## References

- [1] Pedro L Alonso, Jahit Sacarlal, John J Aponte, Amanda Leach, Eusebio Macete, Jessica Milman, Inacio Mandomando, Bart Spiessens, Caterina Guinovart, Mateu Espasa, et al. Efficacy of the RTS, S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial. *The Lancet*, 364(9443):1411–1420, 2004.
- [2] Per Kragh Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical Models Based on Counting Processes*. Springer Science & Business Media, 2012.
- [3] Per Kragh Andersen and Richard D Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, pages 1100–1120, 1982.
- [4] Norman E Breslow. Contribution to the discussion of the paper by DR Cox. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34(2):216–217, 1972.
- [5] Daniel Chandramohan, Seth Owusu-Agyei, Ilona Carneiro, Timothy Awine, Kwame Amponsa-Achiano, Nathan Mensah, Shabbar Jaffar, Rita Baiden, Abraham Hodgson, Fred Binka, et al. Cluster randomised trial of intermittent preventive treatment for malaria in infants in area of high, seasonal transmission in Ghana. *British Medical Journal*, 331(7519):727–733, 2005.
- [6] Yin Bun Cheung, Ying Xu, Sze Huey Tan, Felicity Cutts, and Paul Miligan. Estimation of intervention effects using first or multiple episodes in clinical trials: The Andersen-Gill model re-examined. *Statistics in Medicine*, 29(3):328–336, 2010.

- [7] L Kathryn Durham, Ira M Longini, M Elizabeth Halloran, John D Clemens, Nizam Azhar, and Malla Rao. Estimation of vaccine efficacy in the presence of waning: application to cholera vaccines. *American Journal of Epidemiology*, 147(10):948–959, 1998.
- [8] Johan Gabrielsson and Daniel Weiner. *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*, volume 1. CRC Press, 2001.
- [9] Felicia P Hardnett and Charles E Rose. Measuring the potential role of frailty in apparent declining efficacy of HIV interventions. *HIV Clinical Trials*, 16(6):219–227, 2015.
- [10] Antje Jahn-Eimermacher, Jean-Baptist du Prel, and Heinz-Josef Schmitt. Assessing vaccine efficacy for the prevention of acute otitis media by pneumococcal vaccination in children: a methodological overview of statistical practice in randomized controlled clinical trials. *Vaccine*, 25(33):6237–6244, 2007.
- [11] Mona N Kanaan and C Paddy Farrington. Estimation of waning vaccine efficacy. *Journal of the American Statistical Association*, 97(458):389–397, 2002.
- [12] Paul C Lambert and Patrick Royston. Further development of flexible parametric models for survival analysis. *Stata Journal*, 9(2):265, 2009.
- [13] Peter A Lewis and Gerald S Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.
- [14] Clara Menendez, Elizeus Kahigwa, Rosmarie Hirt, Penelope Vounatsou, John J Aponte, Fidel Font, Camilo J Acosta, David M Schellenberg, Claudia M Galindo, John Kimario, et al. Randomised placebo-controlled trial of iron supplementation and malaria chemoprophylaxis for prevention of severe anaemia and malaria in Tanzanian infants. *The Lancet*, 350(9081):844–850, 1997.
- [15] W. P. O’Meara and T. Lang. Malaria vaccine trial endpoints - bridging the gaps between trial design, public health and the next generation of vaccines. *Parasite Immunology*, 31:574–581, 2009.
- [16] Edsel A. Peña. Asymptotics for a class of dynamic recurrent event models. *Journal of Nonparametric Statistics*, 28(4):716–735, 2016.



- [17] Edsel A Peña, Elizabeth H Slate, and Juan R González. Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137(6):1727–1747, 2007.
- [18] Edsel A Peña, RL Strawderman, and Myles Hollander. Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, 96(456):1299–1315, 2001.
- [19] Patrick Royston and Mahesh KB Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197, 2002.
- [20] WHO. *Guidelines for monitoring the durability of long-lasting insecticidal mosquito nets under operational conditions*. WHO; Geneva, 2011.
- [21] WHO. *Seasonal malaria chemoprevention with sulfadoxine-pyrimethamine plus amodiaquine in children: a field guide*. WHO; Geneva, 2013.
- [22] WHO. *Background paper on the RTSS/AS01 malaria vaccine*. WHO; Geneva, September 2015.
- [23] Ying Xu, Yin Bun Cheung, Kwok Fai Lam, and Paul Milligan. Estimation of summary protective efficacy using a frailty mixture model for recurrent event time data. *Statistics in Medicine*, 31(29):4023–4039, 2012.
- [24] Ying Xu, Kwok Fai Lam, and Yin Bun Cheung. Estimation of intervention effects using recurrent event time data in the presence of event dependence and a cured fraction. *Statistics in Medicine*, 33(13):2263–2274, 2014.