

# Prediction of Event Time for A Time-to-Event Endpoint Under A Piecewise Exponential Model

Liang S. Chen<sup>1</sup>, Xiaoyu Dong<sup>2</sup>

<sup>1</sup> Pfizer, Inc, 219 E 42nd St., New York, NY 10017

<sup>2</sup>Xiaoyu Dong, Department of Mathematics and Statistics,  
University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250

## Abstract

Randomized clinical trials commonly include one or more planned interim analyses. With a time-to-event endpoint, timing of interim analysis or final analysis is usually event-driven. Because these analyses involve timeline and resource planning, it is worthwhile to predict timing of these analyses early and accurately. Parametric models on observed event data from on-going clinical trials were used in Bagiella and Heitjan (2001) and Ying and Heitjan (2008). However, parametric model may not be flexible enough to fit the real data well. We propose the piecewise exponential (**PE**) model to fit the observed event data and estimate the parameters of the model, then predict timing of interim analysis or final analysis. **PE** model is quite flexible, and can fit most of time-to-event data quite well. Accuracy of prediction is assessed by simulation studies by comparing the performance of exponential (**E**) model and **PE** model.

**Key Words:** time-to-event, predicting analysis time, interim analysis, piecewise exponential model

## 1. Introduction

Randomized clinical trials commonly include one or more planned interim analyses. With a time-to-event endpoint, timing of interim analyses or final analysis is usually event-driven. Because these analyses involve timeline and resource planning, it is worthwhile to predict timing of these analyses early and accurately.

Parametric models on observed event data were discussed by several authors. With observed event data, Bagiella and Heitjan (2001) discussed estimating parameters and making predictions under an exponential (**E**) model. Ying and Heitjan (2008) extended this approach to more general parametric model, Weibull model. However, parametric model may not be flexible enough to fit the real data well. Ying et al (2004) also proposed a non-parametric prediction model.

We propose the piecewise exponential (**PE**) model to fit the observed event data and estimate the parameters of the model, then predict timing of interim analysis or final analysis. **PE** model is quite flexible, and can fit most of time-to-event data quite well.

Simulation results are given by comparing **E** mode and **PE** model.

## 2. Statistical Method

### 2.1 Piecewise Exponential (PE) Model

For a defined event, the hazard function of a PE model with  $J$  pieces is given by

$$\lambda(t|\tau_1, \tau_2, \dots, \tau_{J-1}) = \sum_{j=1}^J I_{(\tau_{j-1}, \tau_j]}(t)\lambda_j \tag{1}$$

where  $I_A(t)$  is a indication function, and  $I_A(t) = \begin{cases} 1, & \text{if } t \in A \\ 0, & \text{if } t \notin A \end{cases}$ .  $\tau_0 = 0$  and  $\tau_J = \infty$ ,  $\lambda_j > 0$ , for  $j = 1, \dots, J$  and  $0 < \tau_1 < \tau_2 < \dots < \tau_{J-1} < \infty$ . The survival function will be

$$S(t|\tau_1, \tau_2, \dots, \tau_{J-1}) = e^{-\int_0^t \lambda(v|\tau_1, \tau_2, \dots, \tau_{J-1})dv} \tag{2}$$

Let  $f(t)$  be the probability density function, then  $f(t)$  can be written as

$$\begin{aligned} f(t|\tau_1, \tau_2, \dots, \tau_{J-1}) &= S(t|\tau_1, \tau_2, \dots, \tau_{J-1})h(t|\tau_1, \tau_2, \dots, \tau_{J-1}) \\ &= S(t|\tau_1, \tau_2, \dots, \tau_{J-1}) \times \sum_{j=1}^J I_{(\tau_{j-1}, \tau_j]}(t)\lambda_j \end{aligned} \tag{3}$$

It can be shown that  $\int_0^\infty f(t|\tau_1, \tau_2, \dots, \tau_{J-1})dt = 1$

### 2.2 Likelihood Function for PE Model with Dropouts

Let  $T$  and  $V$  be the independent random variables of the underlying time-to-event and time-to-dropout.  $c$  is denoted as the censored time (administrative censoring only).  $Y$  is the observed survival time. For  $i^{\text{th}}$  subject, the observed survival time is

$$Y_i = \begin{cases} T_i, & \text{if it is an event} \\ V_i, & \text{if it is a dropout} \\ c_i, & \text{if it is censored} \end{cases} \text{ and } Y_i = \min(T_i, V_i, c_i)$$

Let  $d_i$  represent the event indicator for subjects with  $d_i = 1$  if  $Y_i = T_i$  and  $d_i = 0$  if  $Y_i = V_i$  or  $c_i$ ; let  $l_i$  represent the dropout indicator for subjects with  $l_i = 1$  if  $Y_i = V_i$  and  $l_i = 0$  if  $Y_i = T_i$  or  $c_i$ .

Under PE model, assume that  $Y \sim \text{PE}(\lambda)$  for time-to-event and  $Y \sim \text{PE}(\nu)$  for time-to-dropout, where  $\lambda$  is a  $J$  dimension vector and  $\nu$  is a  $K$  dimension vector. With given nodes  $\tau_1, \tau_2, \dots, \tau_{J-1}$  for time-to-event and  $\sigma_1, \sigma_2, \dots, \sigma_{K-1}$  for time-to-dropout, the probability density functions of  $Y$  for time-to-event and time-to-dropout are denoted by  $f(y)$  and  $g(y)$ , and the corresponding distribution functions are denoted by  $F(y)$  and  $G(y)$ . Then the likelihood function for  $n$  subjects is expressed as follows

$$L = \prod_{i=1}^n [f(y_i, \lambda)]^{d_i} [1 - F(y_i, \lambda)]^{1-d_i} [g(y_i, \nu)]^{l_i} [1 - G(y_i, \nu)]^{1-l_i} \tag{4}$$

### 2.3 MLE of $\lambda_j$ with Known Node Positions

From (4) we can see that the likelihood functions for time-to-event and time-to-dropout are well separated, thus we only need to use the corresponding part to estimate the parameters. For  $\lambda$ , consider

$$\begin{aligned}
 L(\lambda) &= [f(y_i, \lambda)]^{d_i} [1 - F(y_i, \lambda)]^{1-d_i} \\
 &= \prod_{i=1}^n \{h(t_i)[1 - F(\lambda, t_i)]\}^{d_i} [1 - F(\lambda, t_i)]^{1-d_i} \\
 &= \prod_{i=1}^n [h(t_i)]^{d_i} [1 - F(\lambda, t_i)]
 \end{aligned} \tag{5}$$

For a simple case with three pieces, the parameters of interest are  $\lambda_1, \lambda_2, \lambda_3$ . Let  $n_j$  be the number of subjects with observed time-to-event  $y_{j,i}$  in  $(\tau_{j-1}, \tau_j]$ ,  $j = 1, 2$  and 3 with  $\tau_0 = 0$  and  $\tau_3 = \infty$ , and  $n_1 + n_2 + n_3 = n$ . The event indicators in each of three pieces are  $d_{j1}, d_{j2}, \dots, d_{jn_j}$ ,  $j = 1, \dots, 3$ . Then the likelihood  $L$  can be partition into three parts,

$$\begin{aligned}
 L &= \prod_{j=1}^3 \prod_{i=1}^{n_j} [h(t_{ji})]^{d_{ji}} [1 - F(t_{ji}, \lambda)] \\
 &= \left\{ \prod_{i=1}^{n_1} \lambda_1^{d_{1i}} e^{-\lambda_1 t_{1i}} \right\} \times \left\{ \prod_{i=1}^{n_2} \lambda_2^{d_{2i}} e^{-\lambda_1 \tau_1} e^{-\lambda_2 (t_{2i} - \tau_1)} \right\} \\
 &\quad \times \left\{ \prod_{i=1}^{n_3} \lambda_3^{d_{3i}} e^{-\lambda_1 \tau_1} e^{-\lambda_2 (\tau_2 - \tau_1)} e^{-\lambda_3 (t_{3i} - \tau_2)} \right\} \\
 &= \{(\lambda_1)^{\sum_{i=1}^{n_1} d_{1i}} e^{-\lambda_1 \sum_{i=1}^{n_1} t_{1i}}\} \times \{(\lambda_2)^{\sum_{i=1}^{n_2} d_{2i}} e^{-n_2 \lambda_1 \tau_1} e^{-\lambda_2 \sum_{i=1}^{n_2} (t_{2i} - \tau_1)}\} \\
 &\quad \times \{(\lambda_3)^{\sum_{i=1}^{n_3} d_{3i}} e^{-n_3 \lambda_1 \tau_1} e^{-n_3 \lambda_2 (\tau_2 - \tau_1)} e^{-\lambda_3 \sum_{i=1}^{n_3} (t_{3i} - \tau_2)}\} \\
 &= (\lambda_1)^{\sum_{i=1}^{n_1} d_{1i}} e^{-\lambda_1 \{[\sum_{i=1}^{n_1} t_{1i}] + (n_2 + n_3) \tau_1\}} \times (\lambda_2)^{\sum_{i=1}^{n_2} d_{2i}} e^{-\lambda_2 \{[\sum_{i=1}^{n_2} (t_{2i} - \tau_1)] + n_3 (\tau_2 - \tau_1)\}} \\
 &\quad \times (\lambda_3)^{\sum_{i=1}^{n_3} d_{3i}} e^{-\lambda_3 \sum_{i=1}^{n_3} (t_{3i} - \tau_2)}
 \end{aligned} \tag{6}$$

Then,

$$\log L(\lambda_1) = \left( \sum_{i=1}^{n_1} d_{1i} \right) \times \log(\lambda_1) - \lambda_1 \left[ \sum_{i=1}^{n_1} t_{1i} + (n_2 + n_3) \tau_1 \right]$$

It's easy to show

$$\begin{aligned}
 \frac{\partial}{\partial \lambda_1} [\log L(\lambda_1)] &= 0 \implies \hat{\lambda}_1 = \frac{\sum_{i=1}^{n_1} d_{1i}}{\sum_{i=1}^{n_1} t_{1i} + (n_2 + n_3) \tau_1} \\
 &= \frac{\text{\# of events in } (0, \tau_1]}{\text{Total follow up time in } (0, \tau_1]}
 \end{aligned}$$

The general formula for MLE of  $\lambda_j$  is

$$\hat{\lambda}_j = \begin{cases} \frac{\sum_{i=1}^{n_j} d_{ji}}{\sum_{i=1}^{n_j} (t_{ji} - \tau_{j-1}) + (\tau_j - \tau_{j-1}) \sum_{p=j+1}^J n_p}, & \text{when } 1 \leq j < J \\ \frac{\sum_{i=1}^{n_j} d_{ji}}{\sum_{i=1}^{n_j} (t_{ji} - \tau_{j-1})}, & \text{when } j = J \end{cases} \tag{7}$$

where  $\sum_{i=1}^{n_j}$  is the summation over all subjects whose event time is between  $\tau_{j-1}$  and  $\tau_j$ . For the drop-out rate, the result is similar as (7).

## 2.4 Node Selection

Optimal nodes can be selected by maximizing the likelihood function when number of nodes are fixed.

## 2.5 Prediction of Event Number

After we obtain the MLE of parameters based on the observed data at  $t_0$ , we can predict the number of events at a future time  $t$ . Following the notation from Bagiella and Heitjan (2001), let the starting time of the study be the time origin, then define

$$T = e + Y_{obs} \quad (8)$$

where  $e$  is the interval length between origin and the enrollment time,  $Y_{obs}$  is the observed length for time to event. The predicted number of events at  $t$ ,  $ED(t_0, t)$ , can be evaluated by

$$ED(t_0, t) = D(t_0) + Q(t_0, t) + R(t_0, t) \quad (9)$$

Where  $D(t_0)$  is the observed number of events at or before  $t_0$ ,  $Q(t_0, t)$  is the expected number of events between  $t_0$  and  $t$  among those censored at or before  $t_0$ , and  $R(t_0, t)$  is the expected number of events between  $t_0$  and  $t$  among those to be enrolled between  $t_0$  and  $t$ .

More specifically,

$$\begin{aligned} Q(t_0, t) &= \sum_{i=1}^{NQ} P_{i, d_i=1}(Y_i + e_i \leq t | Y_i + e_i > t_0) \\ &= \sum_{i=1}^{NQ} \frac{p_{i, d_i=1}(t_0 - e_i < Y_i \leq t - e_i)}{p_i(Y_i > t_0 - e_i)} \\ &= \sum_{i=1}^{NQ} \frac{\int_{t_0 - e_i}^{t - e_i} f(y_i, \lambda) [1 - G(y_i, \nu)] dy_i}{[1 - F(t_0 - e_i, \lambda)] [1 - G(t_0 - e_i, \nu)]} \end{aligned} \quad (10)$$

where  $NQ$  is the number of subjects who are censored at  $t_0$ . Given the enrollment times are known for all the subjects,  $R(t_0, t)$  is computed by

$$\begin{aligned} R(t_0, t) &= \sum_{i=1}^{NR} P_{i, d_i=1}(e_i < Y_i + e_i \leq t) \\ &= \sum_{i=1}^{NR} \int_0^{t - e_i} f(y_i, \lambda) [1 - G(y_i, \nu)] dy_i \end{aligned} \quad (11)$$

where  $NR$  is the number of subjects to be enrolled between  $t_0$  and  $t$ .

## 3. Example

Our model is illustrated by an example below.

This is a randomized, phase 3 trial for the treatment of a solid tumor cancer comparing an experimental treatment with a control treatment. Below are the main bullet points in trial design:

- Subjects are randomized at 1:1 ratio to receive either experimental treatment or control treatment.
- The primary endpoint is Progression Free Survival (**PFS**).
- The enrollment rate is 1 subject/day.
- N=600 (480 **PFS** events are required).
- There is 1 interim analysis at 50% of information fractions (240 **PFS** events).

The goal is to predict the time for interim analysis based on data when 160 events are observed.

Computer-generated data (complete data) were used for this example. Assume time-to-event and time-to-dropout follow **PE** distributions with distribution function F(t) and G(t) for the combined treatment groups respectively.

$$F(t|\tau_1, \tau_2, \tau_3) = \begin{cases} 1 - e^{-\lambda_1 t}, & \text{if } 0 < t \leq \tau_1 \\ 1 - e^{-\lambda_1 \tau_1} * e^{-\lambda_2 * (t - \tau_1)}, & \text{if } \tau_1 < t \leq \tau_2 \\ 1 - e^{-\lambda_1 \tau_1} * e^{-\lambda_2 * (\tau_2 - \tau_1)} * e^{-\lambda_3 * (t - \tau_2)}, & \text{if } \tau_2 < t \leq \tau_3 \\ 1 - e^{-\lambda_1 \tau_1} * e^{-\lambda_2 * (\tau_2 - \tau_1)} * e^{-\lambda_3 * (\tau_3 - \tau_2)} * e^{-\lambda_4 * (t - \tau_3)}, & \text{if } t > \tau_3 \end{cases}$$

where  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0.00080, 0.02200, 0.00060, 0.00300)$  and  $(\tau_1, \tau_2, \tau_3) = (35, 46, 170)$  were selected arbitrarily.

$$G(t|\sigma_1, \sigma_2, \sigma_3) = \begin{cases} 1 - e^{-v_1 t}, & \text{if } 0 < t \leq \sigma_1 \\ 1 - e^{-v_1 \sigma_1} * e^{-v_2 * (t - \sigma_1)}, & \text{if } \sigma_1 < t \leq \sigma_2 \\ 1 - e^{-v_1 \sigma_1} * e^{-v_2 * (\sigma_2 - \sigma_1)} * e^{-v_3 * (t - \sigma_2)}, & \text{if } \sigma_2 < t \leq \sigma_3 \\ 1 - e^{-v_1 \sigma_1} * e^{-v_2 * (\sigma_2 - \sigma_1)} * e^{-v_3 * (\sigma_3 - \sigma_2)} * e^{-v_4 * (t - \sigma_3)}, & \text{if } t > \sigma_3 \end{cases}$$

where  $(v_1, v_2, v_3, v_4) = (0.00040, 0.00200, 0.00010, 0.00005)$  and  $(\sigma_1, \sigma_2, \sigma_3) = (40, 70, 270)$  were selected arbitrarily.

After complete data were generated, then we can determine calendar time  $t_0$  (160 event were observed) and observed data up to  $t_0$ . The 1<sup>st</sup> subject was randomized on 01/03/07. By 04/01/08, 453 subjects had been randomized with 160 events and 22 dropouts. For the observed data, the rest of 271 (453-160-22) subjects were censored at  $t_0$ .

Applying the methods in Sections 2.2 and 2.4, it was estimated that

$$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4, \hat{\lambda}_5, \hat{\lambda}_6) = (0.0006662225, 0.02024965, 0.003753351, 0.0003756446, 0.004668930, 0.002440955)$$

, and

$$(\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\tau}_4, \hat{\tau}_5) = (35, 45, 51, 170, 196)$$

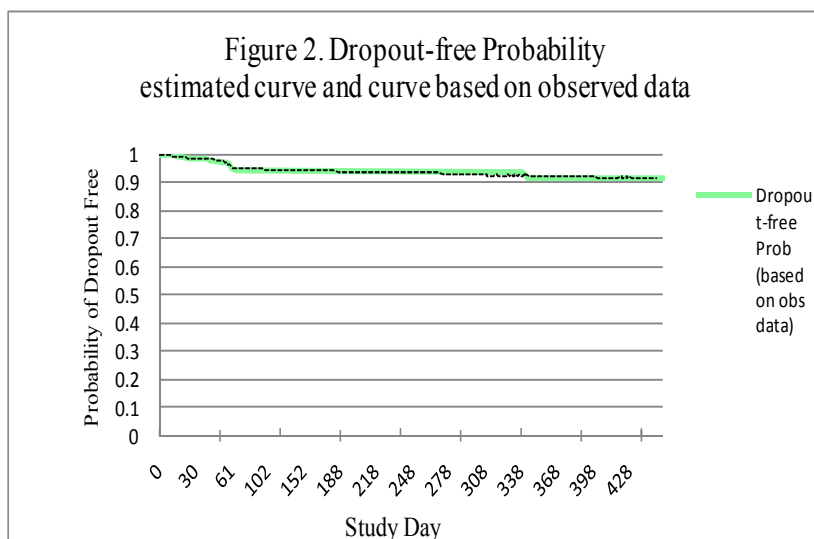
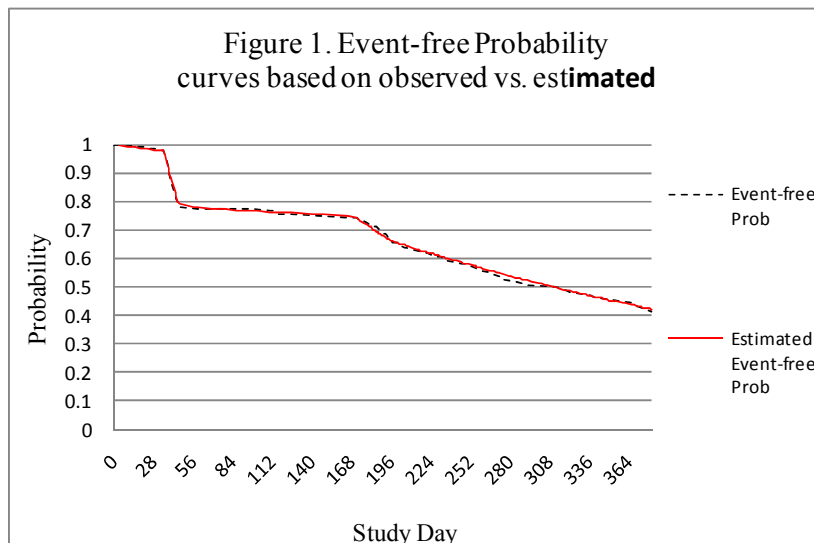
$$(\hat{v}_1, \hat{v}_2, \hat{v}_3, \hat{v}_4) = (0.0006287726, 0.0003043908, 0.003033367, 0.00008777127)$$

and

$$(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3) = (18, 53, 63)$$

The estimates of nodes  $(\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\tau}_4, \hat{\tau}_5)$  and  $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$  are considered optimal if 5 nodes for time to event and 3 nodes for time to dropout are selected, respectively.

Figure 1 and Figure 2 are the figures for observed event-free probability versus estimated event-free probability, and observed dropout-free probability versus estimated dropout-free probability based on observed data up to  $t_0$ .



Results of prediction are given in Table 1. Applying formulas (10) and (11), we predicted that 240 events would occur on 07/31/08. Since we have the complete data, we can compare our prediction with actual cumulative number of events in complete data. In the complete data, 240th events occurred on 07/30/08.

Table 1 Results on Predicted Number of Events

Results	Date for 240 events
Actual from complete data	07/30/08
Estimated from observed data based on PE model	07/31/08

## 4. Simulations

To assess the model performance, we generate the survival data from the Exponential distribution and the Weibull distribution respectively.  $t_0 = 10, 13.3$  and  $16^{\text{th}}$  month from origin are chosen to be the current calendar time with observed data. Number of deaths are predicted at  $T=20, 30,$  and  $40^{\text{th}}$  month from origin under each method. Then compare the prediction accuracy of our **PE** model and the conventional **E** model under each data set using 100 Monte Carlo simulations.

### 4.1 Simulation settings

#### 4.1.1 Generate Data from Exponential Distribution

The *p.d.f.* of Exponential ( $\lambda$ ) is

$$f(t) = \lambda e^{-\lambda t}, \text{ if } t > 0$$

Set the death hazard rates for the control and treatment as  $\lambda_1 = 0.003851$  per day,  $\lambda_2 = 0.002567$  per day, and the hazard rates for the drop out of the two groups are  $\nu_1 = 0.001155$  per day,  $\nu_2 = 0.001155$  per day. Total sample size is 600 with equal allocation for the treatment and control group. In addition, we assume the enrollment rate is 1 subject/day until  $600^{\text{th}}$  day. Both complete data and observed data are generated.

#### 4.1.2 Generate Data from the Weibull Distribution

The simulation settings are very similar as 4.1.1, expect the true data are generated under the Weibull distribution. The *p.d.f.* of Weibull( $a,b$ ) is

$$f(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}, x > 0$$

where  $a > 0$  is the shape parameter, and  $b > 0$  is the scale parameter.

#### 4.1.3 Comparison of E model and PE Model

In **E** model we assumed that time to death and time to dropout followed an exponential distribution for each treatment arm. Parameters in **E** model were estimated following the paper by Donovan et al (2006). In **PE** model we assumed that time to death and time dropout followed a **PE** model for the combined treatment arms. We applied the **PE** model with two nodes for time to death and time to dropout. Then prediction accuracy of **PE** model was compared with the one from **E** model.

### 4.2. Simulation Results

#### 4.2.1. Data from the Exponential Distribution

Results from 100 simulations are listed below.

Table 2 Mean death number and dropout number at current calendar time  $t_0$  and the study end time  $T$ .

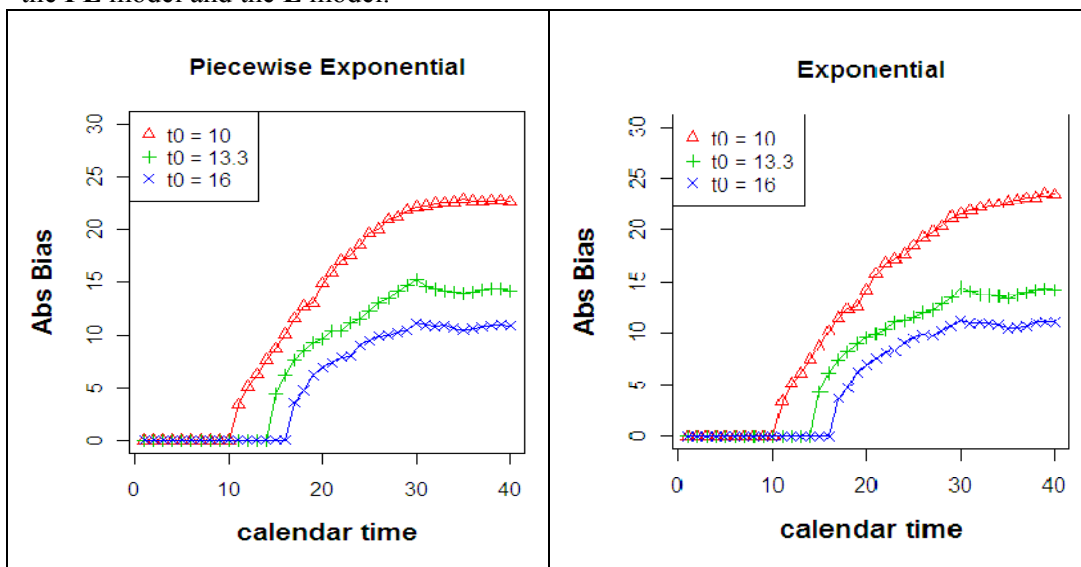
	$t_0 = 10 \text{ m}$	$t_0 = 13.3\text{m}$	$t_0 = 16 \text{ m}$	$T = 40 \text{ m}$
<b>Mean Deaths</b>	95	152	202	424
<b>Mean Dropouts</b>	34	56	74	158

Table 3 Mean and the standard deviation (SD) of the absolute difference between the observed event number and the predicted event number for the future time T using the data at  $t_0$

		$T = 20\text{ m}$		$T = 30\text{ m}$		$T = 40\text{ m}$	
		Mean	SD	Mean	SD	Mean	SD
$t_0 = 10\text{ m}$	<b>PE</b>	14.9	10.0	22.1	14.6	22.7	15.6
	<b>E</b>	14.2	11.1	21.5	18.2	23.5	20.3
$t_0 = 13.3\text{m}$	<b>PE</b>	9.6	7.7	15.3	11.0	14.2	11.8
	<b>E</b>	9.6	7.1	14.5	11.5	14.1	11.7
$t_0 = 16\text{ m}$	<b>PE</b>	6.9	5.2	11.1	8.2	10.9	8.4
	<b>E</b>	6.9	5.2	11.2	8.6	11.1	10.3

Note: **PE** represents the results from the piecewise exponential model with 2 nodes, and **E** represents the results from the exponential model.

Figure 3 Plot of Mean absolute difference between the observed event number in the complete data and the predicted event number for the future time T using the data at  $t_0$  by the **PE** model and the **E** model.



Note: the unit for calendar time is month.

As shown in Table 3, when both data of time to event and time to drop out are from the exponential distribution, the performances of the **PE** and **E** models are very similar. The mean and standard deviation for absolute differences between the observed and predicted event numbers are very close for all scenarios of the current calendar time with observed data  $t_0$  and the future time T for both models.

#### 4.2.2. Data from the Weibull Distribution

Results from 100 simulations are listed below.

Table 4 Mean death number and dropout number at current calendar time  $t_0$  and the study end time T

	$t_0 = 10\text{ m}$	$t_0 = 13.3\text{m}$	$t_0 = 16\text{ m}$	$t_0 = 40\text{ m}$
<b>Mean Death</b>	44	110	169	446



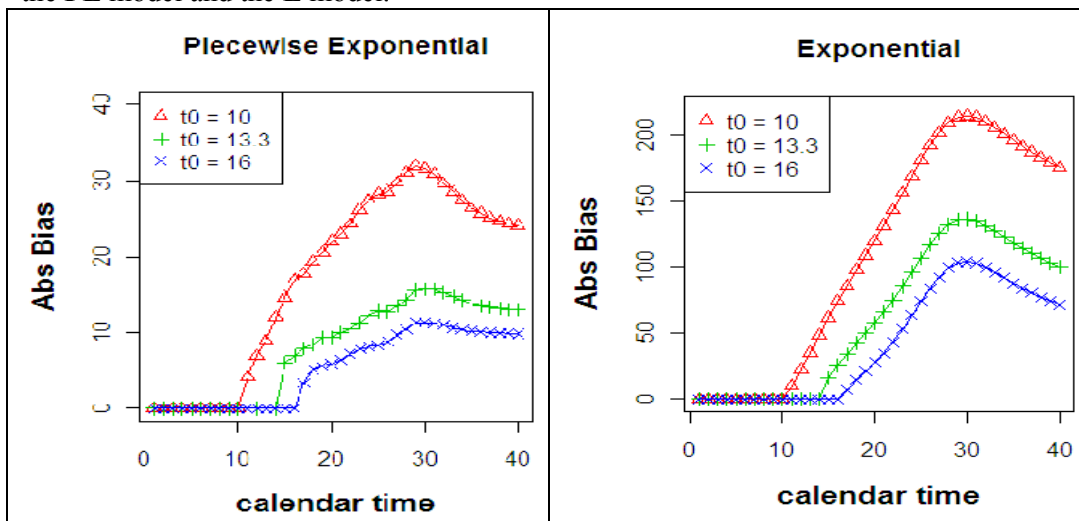
<b>Mean DropOut</b>	44	70	90	155
---------------------	----	----	----	-----

Table 5 Mean and the standard deviation (SD) of the absolute difference between the observed event number and the predicted event number for the future time T using the data at  $t_0$ .

		$T = 20$ m		$T = 30$ m		$T = 40$ m	
		Mean	SD	Mean	SD	Mean	SD
$t_0 = 10$ m	<b>PE</b>	22	16.7	32	24.8	25	19.3
	<b>E</b>	120	14.8	214	20.8	175	21.9
$t_0 = 13.3$ m	<b>PE</b>	10	7.5	16	11.1	13	8.8
	<b>E</b>	58	11.6	136	15.1	100	15.3
$t_0 = 16$ m	<b>PE</b>	6	3.9	12	8.7	10	7.1
	<b>E</b>	28	7.1	104	11.4	72	11.6

Note: **PE** represents the results from the piecewise exponential model with 2 nodes, and **E** represents the results from the exponential model.

Figure 4. Plot of Mean absolute difference between the observed event number in the complete data and the predicted event number for the future time T using the data at  $t_0$  by the **PE** model and the **E** model.



Note: the unit for calendar time is month.

Unlike the simulation results from the previous section, when both data of time to event and time to drop out are from the Weibull distribution, the prediction of the **PE** model is much better than the **E** model based on the results in Table 5. The mean absolute differences between the observed and predicted event numbers from the **E** model is about 6 to 9 times as that from the **PE** model. In general, the **PE** model has smaller standard deviation than that from the **E** model.

#### 4.3 Conclusion from Simulations

In summary, the simulations show that when data is from exponential distribution, the results from the **PE** model are very close to the results from **E** model; when data is from the Weibull distribution (non-exponential distribution), the predictions from **PE** model are much better than those from the **E** model.

## 5. Conclusion/Discussion

PE model is quite flexible to predict number of events for data from unknown distribution. With maximum likelihood method to estimate parameters in PE model, the calculation has the close form, so it is also relative simple.

One of the critical assumptions in this method is to assume future data to be similar to observed data. This means future data within the maximum length of follow up time in the observed data [ $\max_{i^{\text{th}} \text{ subject enrolled on or before } t_0} (t_0 - e_i)$ , where  $e_i$  is the interval length between origin and the enrollment time for  $i^{\text{th}}$  subject] to be similar to the observed data. This also means that the longer follow up data [longer than the maximum length of follow up time in the observed data] to be similar to the last piece in the **PE** model.

## Reference

- Bagiella, Emilia and Heitjan, Daniel F. 2001. Predicting analysis times in randomized clinical trials. *Statistics in Medicine*, 20, 2055-2063
- Donovan, J. Mark, Elliott, Michael R., Heitjan, Daniel F. 2006. Predicting event times in clinical trials when treatment arm is masked. *Journal of Biopharmaceutical Statistics*, 16, 343-356
- Ying, G. S., Heitjan, D. F., Chen, T. T. 2004. Nonparametric prediction of event times in randomized clinical trials. *Clin. Trials* 1:352–361.
- Ying, Gui-shuang and Heitjan, Daniel F. 2008. Weibull prediction of event times in clinical trials. *Pharmaceutical Statistics*, 7, 107-120