Mathematics and Statistics 10(3): 554-561, 2022
DOI: 10.13189/ms.2022.100311 http://www.hrpub.org

Parameter Estimation for Additive Hazard Model Recurrent Event Using Counting Process Approach

Triastuti Wuryandari[1,2], **Gunardi**[1,*], **Danardono**[1]

[1]Departemen of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia

[2]Departemen of Statistics, Faculty of Sains and Mathematics, Universitas Diponegoro, Semarang, Indonesia

Received February 7, 2022; Revised March 29, 2022; Accepted April 15, 2022

Cite This Paper in the following Citation Styles

***(a)**: [1] STriastuti Wuryandari, Gunardi, Danardono, "Parameter Estimation for Additive Hazard Model Recurrent Event Using Counting Process Approach," Mathematics and Statistics, Vol.10, No.3, pp. 554-561, 2022. DOI: 10.13189/ms.2022.100311*

***(b)**: Triastuti Wuryandari, Gunardi, Danardono, (2022). Parameter Estimation for Additive Hazard Model Recurrent Event Using Counting Process Approach. Mathematics and Statistics, 10(3), 554-561, DOI: 10.13189/ms.2022.100311*

Abstract The Cox regression model is widely used for survival data analysis. The Cox model requires a proportional hazard. If the proportional hazard assumption is doubfult, then the additive hazard model can be used, where the covariates act in an additively to the baseline hazard function. If the observed survival time is more than once for one individual during the observation, it is called a recurrent event. The additive hazard model measures risk difference to the effect of a covariate in absolutely, while the proportional hazards model measure hazard ratio in relatively. The risk coefficients estimation in the additive hazard model mimics the multiplicative hazard model, using partial likelihood methods. The derivation of these estimators, outlined in the technical notes, is based on the counting process approach. The counting process approach was first developed by Aalen on 1975 which combines elements of stochastic integration, martingale theory and counting process theory. The method is applied to study about the effect of supplementation on infant growth and development. Based on the processing results, the factors that affect the growth and development of the infant are gender, treatment and mother's education.

Keywords Survival Data, Proportional Hazard, Additive Hazard, Recurrent Event, Counting Process

1 Introduction

Survival analysis is a branch of statistics for analyzing the expected duration of time until one event occurs. Survival data (time to event) is the length of time until an event occurs. Survival data is the realization of a random survival variable, namely a non-negative random variable, T, which is the basis for the formation of models and methods in survival data analysis. If the survival time is influenced by other factors then it can be modeled with a regression model. The regression model for survival data that is widely used is the Cox multiplicative model. The Cox multiplicative model requires proportionality assumption. If the proportional hazard assumption is doubtful, then additive hazard model is alternative in which the covariates act additive manner on an unknown baseline hazard [1].

The relationship between risk factor and time of occurence was described by the multiplicative and additive hazard models. The additive model measures risk differnce to the effect of a covariate in absolutely while the proportional hazards model, measures hazard ratio in relatively [2]. It is quite possible that the relative risk is constant over time, but that the additive risk time varying. In fact, changes in absolute risk with time give no information on changes in relative risk with time. In the additive hazard model, the covariates act additive manner on the baseline hazard and are allowed to time varying. The model with varying coefficients was actually developed first, in 1989 by Aalen, and the model with fixed covariates was developed by Lin and Ying in 1995. For Aalen's model, estimation of the risk coefficient functions can be done using least-squares. Estimator of the risk coefficient is based on the least squares technique. Derivation of its estimator, is based on the counting process approach and similar to the Nelson-Aalen estimator. Hypothesis testing is based on the stochastic integral of the resulting estimator [2]. Lin and Ying's mimic score equation from multiplicative hazard model to estimate risk coefficient in the additive hazard model [3]. The estimation of the regression coefficient in multiplicative hazard model will be obtained through the partial likelihood [2]. The regression coefficient on Lin and Ying's additive hazard model is constant.

Survival analysis is often used to determine the association of risk factors with clinical event such as disease recurrence. Often an event that occurs more than once on an object in a period of observation is called a recurrent event [4]. Recurrent events are often found in the fields of health, epidemiology, insurance, demography, etc. Recurrent events are included in the multivariate failure event because of the correlation between the recurrence events [5]. The survival data are sometimes encountered recurrent event data. Data recurrence if each individual is experiencing the same event repeatedly. Some of the writings are about the frequent instances among others (Prentice, et.al , 1981; Andersen and Gill, 1982; Wei, et.al 1989; Lee et.al 1992; Kelly and Lim, 2000; Lim and Liu, 2007; Lim and Zhang, 2011) [6-11]. They discussed a recurrent event data in the multiplicative hazard model. In the multiplicative hazard model, proportional hazard must be assumed. Anderson and Gill analyzed recurrent events with counting process approach while Prentice, Wei and Peterson used conditional. Kelly and Lim applied a recurrent event multiplicative model to pediatric infectious diseases. Lim and Liu applied the recurrence analysis method to emergency department visits of war pediatric firearm victims while Lim and Zhang applied a multiplicative and additive model to emergency department visits for pediatric firearm victims in Milwauke. If the assumption of proportional hazard is doubtful then the alternative is an additive hazard model. The Cox model emphasizes hazard ratio, while the additive hazard model emphasizes parameter estimation so that the effect of the covariate over time is paid more attention. The additive hazard for recurrent event was presented by Yin and Cai 2004, Schaubel, et.al 2006, Sun et.al 2006 and Zeng and Cai 2010 [12-15]

2 Materials and Methods

The additive hazard model is one of the regression models for survival data. The additive hazard model measures risk difference in absolutely, while the multiplicative hazard model measures hazard ratio in relatively. In this article, we will discuss the estimation of baseline hazard cumulative, estimation of regression coefficients, and an application additive hazard model recurrent event about the effect of giving supplements to infant on their growth and development. .

2.1 The Additive Hazard Model

The additive hazard model is an alternative to the multiplicative hazard model known as the cox model. There are two additive hazard models, namely the Aalen's Additive Hazard Model and Lin and Ying's Additive Hazard Model [2]. The Aalen's Additive Hazard Model was introduced by Aalen in 1989 with a time dependent regression coefficient. The second model is the Lin and Ying's Additive Hazard Model introduced by Lin and Ying in 1994 with a constant regression coefficient [2]. In this paper, we used Lin Ying's additive hazard model, beacause in this model, we provide a closed form estimator for the regression parameters [3]

Let T_i be the time for the i^{th} subject, and C_i be the corresponding censoring time. $X_i = min(T_i, C_i)$, where X_i whose distribution depends on a vector of possibly time-dependent covariates $Z(t) = [Z_1(t), ..., Z_p(t)]$. The data consists of a sample $[T_i, \delta_i, Z_i(t)]$, i=1,..., n, δ_i is the event indicator and $Z_i(t) = [Z_{i1}(t), ..., Z_{ip}(t)]$ is a p vector of possible time dependent covariate. We assume that the hazard rate at time t, for an individual with covariate vector $Z(t)$, is a linear combination of the $Z_j(t)$'s, that is,

$$\lambda(t) = \lambda_0(t) + \sum_{j=1}^{p} \beta_j Z_j(t) \tag{1}$$

where $\lambda_0(t)$ is baseline hazard function. Baseline hazard is the hazard rate for individuals with zero values for all covariates. For continuous covariate, baseline hazard is obtained when the covariates are centered on the mean. The $\beta_j(t)$ are covariate functions estimated from the data.

The estimation of the regression coefficients of the Lin and Ying's Additive Hazard Model used a method similar with score equation on proportional hazard model. The Score equation is obtained from the derivative of the log partial likelihood function. The regression coefficient in Lin and Ying's Additive Hazard model is time independent [2].

For the i^{th} individu, we define

$$Y_i(t) = \begin{cases} 1 & if\ individu\ i^{th}\ is\ at\ risk\ at\ t \\ 0 & if\ individu\ i^{th}\ isn't\ at\ risk\ at\ t \end{cases}$$

To find the estimated regression coefficient in the Lin and Ying's Additive Hazard model, the first step that needs to be done is to estimate the cumulative baseline hazard function. The estimator of the cumulative baseline hazard function can be found using the counting process theory. The counting process, $N_i(t)$, can be described as follows

$$N_i(t) = M_i(t) + \int_0^t Y_i(t)[\lambda_0(t) + \beta^t Z_i(t)]dt \tag{2}$$

where $M_i(t)$ is a martingale. Equation (2) is derived, we get

$$dN_i(t) = dM_i(t) + Y_i(t)\lambda_0(t)dt + Y_i(t)\beta^t Z_i(t)dt \tag{3}$$

Equation (3) is added and integrated, we get

$$\int_0^t \lambda_0(t)dt = \int_0^t \frac{\sum_{i=1}^n dN_i(t) - Y_i\beta^t Z_i(t)dt}{\sum_{i=1}^n Y_i(t)} - \int_0^t \frac{\sum_{i=1}^t dM_i(t)}{\sum_{i=1}^n Y_i(t)} \tag{4}$$

The term $\int_0^t \frac{\sum_{i=1}^n dM_i(t)}{\sum_{i=1}^n Y_i(t)}$ is called the error component where $E(dM_i(t)) = 0$

The estimation of the cumulative baseline hazard function is as follows

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n (dN_i(t) - Y_i(t)\beta^t Z_i(t)dt}{\sum_{i=1}^n Y_i(t)} \tag{5}$$

The estimation of this cumulative baseline hazard function will be substituted for the score equation that will be obtained in the next step.

The regression coefficient estimation in Lin and Ying's additive hazard model uses a score equation that mimics the score equation in the Cox regression model. The score equation in the Cox model is obtained from the derivative of log partial likelihood

The Cox model's score equation is as follows

$$U(\beta) = \sum_{i=1}^{n} \int_0^t Z_i(t) \Big[dN_i(t) - Y_i(t) exp\big(\beta^t Z_i(t) d\Lambda_0(t)\big) \Big] \tag{6}$$

The Score equation on (6) of the Cox proporsional hazard model, the term $exp\beta^t Z_i(t) d\Lambda_0(t)$ is replaced by $\Lambda_0(t) + \beta^t Z_i(t)$ of the Lin and Ying's Additive Hazard Model. The score equation for the Lin and Ying's additive hazard model is as follows:

$$U(\beta) = \sum_{i=1}^{n} \int_0^\infty Z_i(t) \big[dN_i(t) - Y_i(t) d\Lambda_0(t) - Y_i(t)\beta^t Z_i(t) \big] \tag{7}$$

Equation (5) is subtituted to (7), we get

$$U(\beta) = \sum_{i=1}^{n} \int_0^t Z_i(t) - \left[\frac{\sum_{i=1}^{n} Y_i(t) Z_i(t)}{\sum_{i=1}^{n} Y_i(t)} \right] \left[dN_i(t) - Y_i(t)\beta^t Z_i(t) dt \right] \tag{8}$$

The estimate of the regression coefficient, is obtained by solving the equation $U(\beta) = 0$ we get

$$\widehat{\beta} = \left(\sum_{i=1}^{n} \int_0^t Y_i(t) [Z_i(t) - \overline{Z}(t)][Z_i(t) - \overline{Z}(t)] dt \right)^{-1} \left(\sum_{i=1}^{n} \int_0^\infty [Z_i(t) - \overline{Z}(t)(t) dN_i(t)] \right) \tag{9}$$

where

$$\frac{\sum_{i=1}^{n} Y_i(t) Z_i(t)}{\sum_{i=1}^{n} Y_i(t)} = \overline{Z}(t)$$

The variance matrix $\widehat{\beta}$ is estimated by $A^{-1}CA^{-1}$ where

$$A = \sum_{i=1}^{n} \sum_{j=1}^{p} \big(T_i - T_{i-1}\big) \big(Z_i - \overline{Z}(T_i)\big)^t \big(Z_i - \overline{Z}(T_i)\big) \tag{10}$$

$$C = \sum_{i=1}^{n} \delta_i \big(Z_i - \overline{Z}(T_i)\big)^t \big(Z_i - \overline{Z}(T_i)\big) \tag{11}$$

2.2 Recurrent Event

Often an individual experiences the event more than once during the observation period. These events are known as recurrent event [4]. Examples of recurring events include tumor recurrences, epileptic seizures, asthma attacks, migraines, infectious episodes, myocardial infractions, injuries, and admissions to the hospital [5].

Recurrent event is a recurring event of the same type of event from time to time, while the process that produces the data is called a recurrent event process. In recent years, recurrent data has increased in many areas such as public health, business and industry, reliability, social sciences, and insurance. Recurrence event occurs frequently in medical studies, where information is often available on many individuals, each of whom may experience a recurrent transient clinical event during the observation period. Examples include the occurrence of asthma attacks in respiratory studies, epileptic seizures and facts in osteoporosis studies. In business examples include filling out warranty claims on cars or insurance claims for policyholders [16].

The data consists of a sample $[T_{ik}, \delta_{ik}, Z_{ik}(t)]$, $i = 1, ..., n$, where T_{ik} is study time, δ_{ik} is the event indicator and $Z_{ik}(t) = [Z_{ik1}, ..., Z_{ikp}]$ is a p-vector of possibly time dependent covariate. Let T_{ik} be the time when the k^{th} failure occurs for the i^{th} subject and C_{ik} be the corresponding censoring time. T_{ik} is measured from the subject enrollment and the censoring C_{ik} occurs after the subject has been entered into a study to the right of the last known failure time and it is right censoring. When T_{ik} is subject to right censoring, the k^{th} failure time $X_{ik} = min(T_{ik}, C_{ik})$. $X_{ik} = T_{ik}$ if the subject was observed and $X_{ik}(t) = C_{ik}$ if the subject was censored. Let $\delta_{ik} = I(T_{ik} \leq C_{ik}$, where $I(.)$ is an indicator function and takes the value 1 when $T_{ik} \leq C_{ik}$ and is 0 otherwise

In 1994, Lin and Yang introduced the additive hazard model where the regression coefficient is constant. This model is known as the semiparametric additive hazard model [2]. The Lin and Ying's additive hazard model for recurrent event with covariate vector is

$$\lambda_{ik}(t) = \lambda_{0k}(t) + \beta^t Z_{ik}(t) \tag{12}$$

Where $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_p]$ is regression coefficient, λ_{0k} is a baseline hazard function. $\boldsymbol{Z}_{ik}(t) = [Z_{ik1}(t), ..., Z_{ikp}(t)]$ is kxp vector covariates.

Data consists of a sample $[T_{ik}, \delta_{ik}, Z_{ik}]$, i= 1,2,...n and k=1, 2, ..., K where $T_{ik}\rho$ is the event time for i^{th} subject and k^{th} recurrence, δ_{ik} is the event indicator and is a Kxp-vector covariates. We define $Y_{ik}(t) = 1$ if i^{th} subject and k^{th} recurrence that is at risk in t and $Y_{ik}(t) = 0$ if i^{th} subject and k^{th} recurrence that is not risk in t. Illustration of recurrent event observation is presented in Figure 1.

2.3 Counting Process

Counting Process Approach was first developed by Aalen on 1975 who combined elements of stochastic integration, continuous time martingale theory and counting process theory. Counting Process was defined $N_{ik}(t) = 1$; if $T_{ik} \leq t$ and $\delta_{ik} = 1$ and $N_{ik}(t) = 0$, otherwise. Counting Process with the properties that $N_{ik}(0) = 0$, and $N_{ik}(t) \leq \infty$ with probability 1 and the sample path of $n_{ik}(t)$ are right censored sample and piecewise constant with jumps of size 1. The counting process gives us information about when event occurs. We defined the process $\Lambda(t) = \int_0^t \lambda_{ik}(t)dt$, $t \geq 0$, where $\lambda_{0k}(t)$ is called intensity process of the counting process. If the stochastik process $M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(t)\lambda_{ik}(t)dt$, then thus one may think of the counting process as the total number of observed

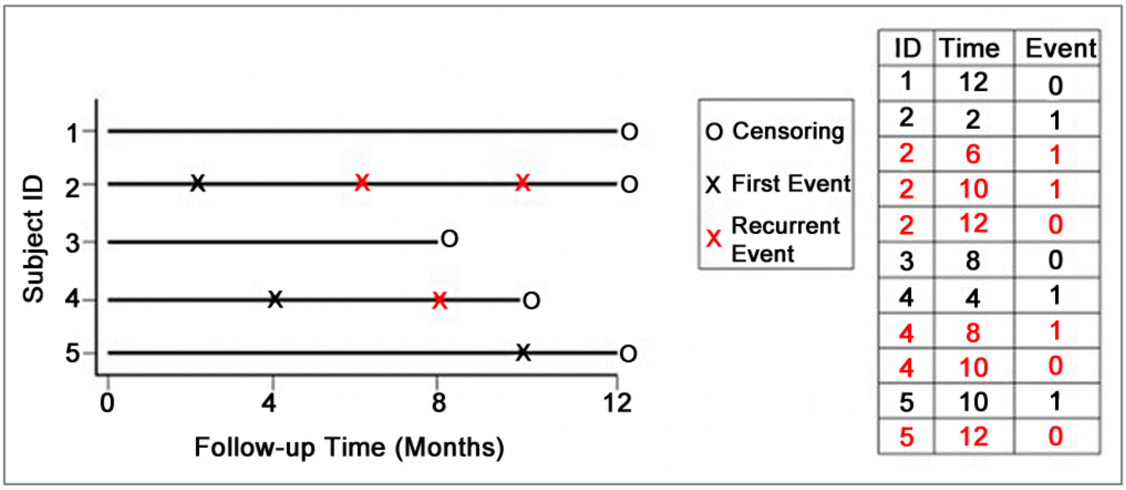

Figure 1. Illustration for Recurrent Event

events and the cumulative intensity process as the total number of expected events up to time t. The difference between these two quantities is a residual like quantity called the counting process martingale [2]. The basic counting process $N_{ik}(t)$ can be decomposed into

$$N_{ik}(t) = M_{ik}(t) + \int_0^t Y_{ik}(t)[\lambda_{0k}(t) + \beta^t Z_{ik}(t)]dt \quad (13)$$

where $M_{ik}(t)$ is martingale.
The general data layout for recurrent data with the counting process approach is presented in Table 1

2.4 Parameter Estimation

There are n subjects and that each subject can experience k failures or recurrent events. Suppose that censoring is non-informative, which means that knowledge of a censoring time for a subject provides no further information about the subject's likelihood of survival at a future time. Let T_{ik} be the time when the k^{th} failure occurs for the i^{th} subject and C_{ik} be the corresponding cesoring time. T_{ik} is measured from the subject's study enrollment and the censoring occurs after the subject has been entered into a study to the right of the last known failure time; thus, it is right censoring. When T_{ik} is subject to right censoring, the k^{th} failure time $X_{ik} = min(T_{ik}; C_{ik})$, i.e., $X_{ik} = T_{ik}$ if the event was observed and $X_{ik} = C_{ik}$ if it is censored. Let $\delta_{ik} = I(T_{ik} \leq C_{ik})$, where $I(.)$ is an indicator function and takes the value 1 when $X_{ik} = T_{ik}$ and 0 otherwise. Let Z_{ik} be a covariate vector of p-dimensions for the i^{th} subject at the k^{th} failure. For each of the K failures, the hazard function for the i^{th} subject with respect to the k^{th} failure. Defined $N_{ik}(t)$ denote the counting process representing the number of event experienced before time t.

2.4.1 Estimation For Cumulative Hazard Baseline

Estimation in the additive hazard model focuses on the cumulative regression function. Estimating the cumulative baseline hazard function uses the counting process theory. The Counting is defined as follows

$$N_{ik}(t) = M_{ik}(t) + \int_0^t Y_{ik}(t)[\lambda_{0k}(t) + \beta^t Z_{ik}(t)]dt \quad (14)$$

The equation (14) is derivatived, we get

$$dN_{ik}(t) = dM_{ik}(t) + Y_{ik}(t)[\lambda_{0k}(t) + \beta^t Z_{ik}(t)]dt \quad (15)$$

The equation (15) is added and integrated, we get estimation of cumulative baseline hazard as follows

$$\widehat{\Lambda}_{0k}(t) = \int_0^t \lambda_{0k} dt \quad (16)$$

The estimation of the cumulative baseline hazard function on (16) will be substituted in the score equation to be obtained in the next step.

2.4.2 Estimation For Regression Coeficient

The regression coefficient of Lin and Ying's model was estimated by using the score equation as in the Cox regression model. The score equation from The Cox regression is obtained by deriving the log partial likelihood function is

$$L(\beta) = \Pi_{k=1}^{K} \Pi_{i=1}^{n} \Pi_{0 \leq t \leq \infty} \left[\frac{Y_{ik}(t) exp(\beta^t Z_{ik}(t))}{\sum_{i=1}^{n} Y_{ik}(t) exp(\beta^t Z_{ik}(t))} \right]^{dN_{ik}(t)} \quad (17)$$

The logarithm of (17) is derived from β to produce the score equation of the multiplicative hazard model. The score equation of the multiplicative hazard model recurrent events is as follows:

$$U(\beta) = \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^t dN_{ik}(t) \Big(Z_{ik}(t) - B \Big) dt \quad (18)$$

where

$$B = -\frac{\sum_{k=1}^{K} \sum_{i=1}^{n} Z_{ik}(t) Y_{ik}(t) exp(\beta^t Z_{ik}(t))}{\sum_{k=1}^{K} \sum_{i=1}^{n} Y_{ik}(t) exp(\beta^t Z_{ik}(t))}$$

Table 1. General Data Layout for Recurrent Event

Subject	Interval	Status	Start	Stop	Z_{i1k}	...	Z_{ipk}
1	1	δ_{11}	t_{110}	t_{111}	Z_{111}	...	Z_{11p}
1	2	δ_{12}	t_{120}	t_{121}	Z_{121}	...	Z_{12p}
.	.	.	.	.	.	...	.
.	.	.	.	.	.	...	.
.	.	.	.	.	.	...	.
1	r_1	δ_{1r_1}	t_{1r_10}	t_{1r_11}	Z_{1r_11}	...	Z_{1r_1p}
2	1	δ_{21}	t_{210}	t_{211}	Z_{211}	...	Z_{21p}
2	2	δ_{22}	t_{220}	t_{221}	Z_{221}	...	Z_{22p}
.	.	.	.	.	.	...	.
.	.	.	.	.	.	...	.
.	.	.	.	.	.	...	.
2	r_2	δ_{2r_2}	t_{2r_20}	t_{2r_21}	Z_{2r_21}	...	Z_{2r_2p}
.	.	.	.	.	.	...	.
n	1	δ_{n1}	t_{n10}	t_{n11}	Z_{n11}	...	Z_{n1p}
n	2	δ_{n2}	t_{n20}	t_{n21}	Z_{n21}	...	Z_{n2p}
.	.	.	.	.	.	...	.
.	.	.	.	.	.	...	.
.	.	.	.	.	.	...	.
n	r_n	δ_{nr_n}	t_{nr_n0}	t_{nr_n1}	Z_{nr_n1}	...	Z_{nr_np}

The equation (18) is decomposed into

$$U(\beta) = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t Z_{ik}(t) - C \tag{19}$$

where

$$C = \big(dN_{ik}(t) - Y_{ik}(t)exp\beta^t Z_{ik}(t)d\widehat{\Lambda}_0(t)\big)dt \tag{20}$$

and

$$d\widehat{\Lambda}_0(t) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{n} dN_{ik}(t)}{\sum_{k=1}^{K}\sum_{i=1}^{n} Y_{ik}(t)exp(\beta^t Z_{ik}(t))}$$

The term $exp(\beta^t Z_{ik}(t)dt$ is rechaged by $\Lambda_{0k}(t) + \beta^t Z_{ik}(t)dt$, we get

$$U(\beta) = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t Z_{ik}(t)[dN_{ik}(t) - Y_{ik}(t)(\lambda_{0k}(t) + \beta^t Z_{ik}(t)dt] \tag{21}$$

Equation (21) is called score equation from additive hazard model recurrent event.

The equation (16) is substituted into (19), we get

$$U(\beta) = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t Z_{ik}(t) - D \tag{22}$$

where

$$D = -\Bigg[\frac{\sum_{k=1}^{K}\sum_{i=1}^{n} Y_{ik}(t)Z_{ik}(t)}{\sum_{k=1}^{K}\sum_{i=1}^{n} Y_{ik}(t)}\Bigg]\big[dN_{ik}(t) - Y_{ik}(t)\beta^t Z_{ik}(t)dt\big] \tag{23}$$

The equation (22) is rewriten as

$$U(\beta) = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t \big[Z_{ik}(t) - \overline{Z}(t)\big]\big[dN_{ik}(t) - Y_{ik}(t)\beta^t Z_{ik}(t)dt\big] \tag{24}$$

where $\overline{Z}_k(t) = \frac{\sum_{i=1}^{n} Y_{ik}(t)Z_{ik}(t)}{\sum_{i=1}^{n} Y_{ik}(t)}$

The estimate of β is obtained by solving the equation $U(\beta) = 0$ and we obtained the consistent estimator β

$$\widehat{\beta} = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t Y_{ik}(t)[Z_{ik}(t) - \overline{Z}_k(t)]^{\otimes 2\,-1}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t [Z_{ik}(t) - \overline{Z}_k(t)]dN_{ik}(t) \tag{25}$$

where $a^{\otimes 2} = aa^t$. The variance covariance matrix of β may be estimated by $A_1^{-1}V_1A_1^{-1}$ where

$$A_1 = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t Z_{ik}(t) - \overline{Z}_k{}^{\otimes 2} Y_{ik}(t)dt$$

and

$$V_1 = \sum_{i=1}^{n}\Big[\sum_{k=1}^{K}\int_0^t Z_{ik}(t) - \overline{Z}_k(t)dN_{ik}(t)$$

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t [Z_{ik}(t) - \overline{Z}_k(t)]dN_{ik}(t)$$

The estimate of β is obtained by solving the equation $U(\beta) = 0$ and we obtained the consistent estimator β

$$\widehat{\beta} = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t Y_{ik}(t)[Z_{ik}(t) - \overline{Z}_k(t)]^{\otimes 2\,-1}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t [Z_{ik}(t) - \overline{Z}_k(t)]dN_{ik}(t) \tag{26}$$

where $a^{\otimes 2} = aa^t$. The variance covariance matrix of β may be estimated by $A^{-1}VA^{-1}$ where

$$A^{-1} = \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t (Z_{ik}(t) - \overline{Z_k})^{\otimes 2} Y_{ik}(t)dt \tag{27}$$

and

$$V=\sum_{i=1}^{n}[\sum_{k=1}^{K}\int_0^t Z_{ik}(t)-\overline{Z_k(t)}dN_{ik}(t)$$
$$\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^t [Z_{ik}(t)-\overline{Z_k(t)}]dN_{ik}(t) \quad (28)$$

2.5 Risk Difference

In the additive hazard model, Risk Difference can be known. Risk Difference is the absolute difference in outcome between a control group and the treatment group, i.e. the percentage of how much the risk of something is reduced if a particular intervention occurs. The Risk Difference is straightforward to interprete which describes the actual difference in the observed risk of event between experimental and control intervention for an individual. It describes the estimated difference in the probability of experiencing the event [17] The Risk Difference was calculated by subtracting the risk for the placebo group with the risk for the treatment group. Risk Difference is defined as

$$\begin{aligned} RD_{ab} &= \lambda_{ij(a)}-\lambda_{ij(b)} \\ &= (\Lambda_{0k}(t)+\beta_{(j)}Z_{ij(a)}(t)-(\Lambda_{0k}(t)+\beta_{(j)}Z_{ij(b)}(t)) \\ &= (\beta_{(j)}Z_{ij(a)}(t)-(\beta_{(j)}Z_{ij(b)}(t) \end{aligned} \quad (29)$$

3 An Application

The case study in this research is data from the Lind's research who discussed iron and zinc deficiency affecting infant's growth and development and increasing infectious disease morbidity during infancy and childhood. Therefore, a combination of iron and zinc supplementation may be a logical prevention strategy [18]. Lind, et.al use analysis of variance to determine the factors that influence infant's growth and development.

The objective of the study was to compare the effects of combined iron and zinc supplementation in infancy with the effects of iron and zinc as single micronutrients on growth, psychomotor, development, and incidence of infectious disease. Design 680 Indonesian infants were randomly assigned to daily supplementation with 10 mg Iron, 10 mg Zinc, 10 mg Iron+10 mg Zinc and placebo from 6 to 12 month of age. Another variable that is thought to have an effect on infant growth and development is the mother's level of education. There are 4 levels of mother's education, 1=edclass no-educ (not graduated from elementary school, 2=edclass6yrs (elementary school), 3=edclass12yrs(secondary school), 4=edclass12more (bachelor's degree) [18].

Data processing used R software with survival package. The regression model used is the additive hazard model for recurrent events. The independent variable and parameter estimation can be seen in Table 2 and Table 3.

According to Table 3, the factors that influence child growth and development are gender, treatment (supplement) and mother's education (edclass).

Risk Different for girl is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{girl}Z_{girl}(t)-(\beta_{boy}Z_{boy}(t) \\ &= (-0.0000637) \end{aligned}$$

It means, girls have a smaller risk of growth and developing disorders than boys

Risk Different for treatment zinc is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{zinc}Z_{zinc}(t)-(\beta_{placebo}Z_{placebo}(t) \\ &= (-0.00699) \end{aligned}$$

It means infant who are given zinc supplements have a smaller risk of growth and developing disorders than infant without supplements

Risk Different for treatment zinc+iron is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{zinc+iron}Z_{zinc+iron}(t)-(\beta_{placebo}Z_{placebo}(t) \\ &= (-0.00984) \end{aligned}$$

It means infant who are given zinc+iron supplements have a smaller risk of growth and developing disorders than infant without supplements

Risk Different for treatment iron is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{iron}Z_{iron}(t)-(\beta_{placebo}Z_{placebo}(t) \\ &= (-0.00592) \end{aligned}$$

It means infant who are given iron supplements have a smaller risk of growth and developing disorders than infant without supplements

Risk Different for mother's education 6yrs is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{6yrs}Z_{6yrs}(t)-(\beta_{no-educ}Z_{no-educ}(t) \\ &= (-0.00529) \end{aligned}$$

It means, infant fron mother's education 6yrs have a smaller risk of growth and developing disorders than infant from mother's education no-educ

Risk Different for infant from mother's education 12yrs is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{12yrs}Z_{12more}(t)-(\beta_{no-educ}Z_{no-educ}(t) \\ &= (-0.000167) \end{aligned}$$

It means, infant from mother's education 12yrs have a smaller risk of growth and developing disorders than infant from mother's education no-educ.

Risk Different for infant from mothe's education 12more is defined as

$$\begin{aligned} RD &= \lambda_{ik(a)}-\lambda_{ik(b)} \\ &= (\beta_{12more}Z_{12more}(t)-(\beta_{no-educ}Z_{no-educ}(t) \\ &= (-0.000337) \end{aligned}$$

Table 2. Independent variables

Variable	categoric
Sex	0=boy; 1=girl
treatment	0=placebo; 1=zinc; 2=zinc+iron; 3=iron
mother's education	0=no-educt; 1=6years; 2=12years; 3=12more

Table 3. Parameter Estimation

Variable	slope	coef	se(Coef)	z	p
Intercept	0.05380	0.0009000	0.0001310	6.88	5.86e-12
sexgirl	-0.00504	-0.0000637	0.0000407	-1.51	0.1170
treatmentzinc	-0.00349	-0.0000699	0.0000577	-1.21	0.2260
treatmentzinc+iron	-0.00507	-0.0000984	0.0000573	-1.72	0.0859
treatmentiron	-0.00337	-0.0000592	0.0000585	-1.01	0.3120
edclass6yrs	-0.00541	-0.0000528	0.0001290	-0.41	0.6820
edclass12yrs	-0.01300	-0.0001670	0.0001290	-1.24	0.1960
edclass12more	-0.01670	-0.0003370	0.0000180	-1.88	0.0604

It means, infant from mother's education 12more have a smaller risk of growth and developing disorders than infant from mother's education no-educ

4 Conclusions

The additive hazard model is an alternative to the regression model for survival data if the proportional hazard assumption is doubtful. The additive hazard regression model measures the additional risk in absolutely while multiplicative hazard regression measures excess risk in relatively. Parameter estimation in the additive hazard model, included estimation of baseline hazard and estimation of regression coefficients carried out using the counting process approach. Based on the case study, the variables that affect the infant growth and development are gender (boy or girls), treatment (supplementation of zinc, iron and a combination of zinc + iron) and mother's education (edclass)

Acknowledgements

We would like to thank the editors and reviewers for their critical reading of the manuscript and the constructive comments that led to great improvement of the earlier version. We would like to thank LPDP, Gadjah Mada University and Diponegoro University for supporting my research.

REFERENCES

[1] Sarker, S., Min, D.K., Black, T. and Lim, H., "Application of Multiplicative and Additive Hazard Models to Injury Prevention among Health care Worker," Advances in Research, vol.5, no.1, pp.1-12., 2015, DOI:10.9734/AIR/2015/18021.

[2] Klein, J.P. and Moeschberger, M.L. "Additive Hazard Models, "in Survival Analysis Techniques Truncated Data, Second Edition, Springer Science, New York, 2003, pp. 329-350.

[3] Lin, D.Y. and Ying, Z., "Semiparametric analysis of the additive risk model, "Biometrika, vol.81, no.1., pp.61-71., 1994, https://doi.org/10.2307/2337050.

[4] Kleinbaum, D.G. and Klein, M. "Recurrent Event Survival Analysis, "in Survival Analysis A Self Learning Text, Second Edition, Springer Science, New York, 2005, pp. 332-371.

[5] Lim, H., and Zhang, X. "Additive and Multiplicative Hazard Modelling for Recurrent event data Analysis, "BMC Medical Research Methodologi. Vol.11 No.101, pp. 1-12, 2011, doi: 10.1186/1471-2288-11-101

[6] Prentice, R., L., Williams, B.J., and Peterson, A..V.,. "On The Regression Analysis of Multivariate Failure Time Data, "Biometrika, Vol.68, No.2, pp.373-379. 1981 https://doi.org/10.1093/biomet/68.2.37

[7] Andersen, P. K. and Gill, R. D. "Cox's regression model for counting processes: a large sample study', "The Annals of Statistics, vol.10, No. 4, pp. 1100 1120, 1982. https://www.jstor.org/stable/2240714.

[8] Wei LJ, Lin DY, Weissfeld L., "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, "Journal of American Statistical Association, Vol. 84, No. 408, pp.1065-1073, 1989, https://doi.org/10.2307/2290084.

[9] Lee E.W., Wei L.J., and Amato D.A., "Cox-type regression analysis for large number of small groups of correlated failure time observations, in Klein, J.P. and Goel, P.K., (eds), Survival Analysis: State of the Art, Kluwer Academic Publisher; 1992, Dordrecht, 1992, pp. 237-247.

[10] Kelly P., and Lim L-Y., "Survival analysis for recurrent event data: an application to childhood infectious diseases, "Statistics in Medicine, vol.19, no.1. pp. 13-33, 2000, DOI:10.1002/(SICI)1097-0258(20000115)19:13.0.CO;2-5

[11] Lim H.J., Liu J. and Melzer-Lange M, "Comparison of methods for analyzing recurrent event data: application to emergency department visits of pediatric firearm victims, "Accident Analysis Prevention, vol. 39, no.2, pp.290-299. 2007, DOI: 10.1016/j.aap.2006.07.009

[12] Yin, G. and Cai, J., "Additive hazards model with multivariate failure time data, "Biometrika, Vol.91, No.4, pp.801-818, 2004, https://www.jstor.org/stable/20441145

[13] Schaubel, D., Zeng, D., and Cai, J., ”A semiparametric Additive Rates Model for Recurrent Event, ”Lifetime Data Anal, Vol.12, No.4, pp.389-406, 2006, doi:10.1007/s10985-006-9017

[14] Zeng, D., and Cai, J., ”A semiparametric additive rate model for recurrent event with an informative terminal event, ”Biometrika, vol.97, no.3. pp. 699-712, 2010, https://www.jstor.org/stable/25734117

[15] . Sun, L., Park, D., Sun, G. ”The additive hazards model for recurrent gap times. ”Statistica Sinica, vol. 16, no. 3, pp. 919-932, https://www.jstor.org/stable/24307581

[16] Cook, R.J., and Lawless, J.F., ”Introduction, ”in The Statistical Analysis of Recurrent Event, tenth Edition, Springer Science, New York, 2007, pp. 1-7.

[17] Zhang, Z., Ambrogi, F., Bokov, A., Gu, H., Beurs,E., and Eskar, K., ”Estimate Risk Difference and Number Needed to Tried in Survival Analysis, ”Annals of Translational Medicine, vol.6, no.7, pp.120-128, 2018, doi: 10.21037/atm2018.0136,

[18] Lind,T., Lönnerdal B., Stenlund, H., Gamayanti, I., Ismail, D., Seswandhana, R., and Persson, L. ”A community-based randomized controlled trial of iron and zinc supplementation in Indonesian infants: effects on growth and development, ”The American Journal of Clinical Nutrition vol.77, no.4, pp.883-890, 2003, doi: 10.1093/ajcn/77.4.883.