An Extended Kaplan-Meier Estimator for Time-to-Event Data with Informative Censoring

Wei Li[1], Misun Lee[1]

[1]Astellas Pharma Inc., 1 Astellas Way, Northbrook, IL 60062

Abstract

In many clinical trials, an important endpoint is rate of treatment success which can be analyzed by crude rate or time-to-event approaches. In trials where patients may discontinue treatment or die due to poor response to treatment, the use of standard survival methods such as the Kaplan-Meier estimator and log-rank test may yield severely biased results. This is because the independence of censoring (or noninformative censoring) assumption is often violated. The analysis of time-to-success must therefore account for informative censoring. We propose a two-step method that is capable of handling multiple censoring mechanisms including informative censoring in a trial. Step one is to utilize observed data to assign a unique likelihood index to each censored patient that quantitatively measures the censored patient's likelihood of achieving success relative to the remaining patients. In Step two, we propose an extended Kaplan-Meier (EKM) estimator that adjusts number of patients "at risk" based on the likelihood index of patient censored at each time point. Unlike standard KM estimator, the EKM is flexible in accounting for various censoring mechanisms and does not ignore other observed data of censored patients. We illustrate the application of EKM using a case study. While developed for time-to-success, the EKM is applicable to general time-to-event endpoints.

Key Words: Informative censoring; Extended Kaplan-Meier estimator; Time-to-event; Likelihood index

1. Introduction

In clinical trials, time to event analyses usually focus on failure time, such as time to death (overall survival) and time to development of a disease (disease free survival and progression free survival). In those analyses, treatment failure and death are the outcome events and patients without the event observed are censored. The Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958) and log-rank test have been the standard methods for failure time analysis for half of a century and have been generally robust in analyzing such data. However, like many statistical methods, they have a limitation in that they require the assumption of independent censoring (also called noninformative censoring). The independent censoring assumption dictates that censoring time is independent of survival time. In other words, censoring (or reason for censoring) is unrelated to patient's response to treatment and patient's underlying disease process.

In some clinical trials, the outcome of interest is treatment success rather than treatment failure. This can be analyzed by a crude rate approach such as objective response rate in oncology studies or a time to event approach. An important difference between time-to-failure analysis and time-to-success analysis is with how to treat failure. In the time to treatment failure (or death), we do not censor patients with observed failure or death. However, in the time to treatment success, patients with treatment failure and death are

usually censored. Analysis of time-to-success presents a unique challenge because independent censoring assumption is often violated as censoring of patients due to treatment failure or death is directly related to patient's response to treatment and is thus informative censoring. Therefore, use of standard survival analysis methods such as KM estimator which rely on the independent censoring assumption may yield severely biased results. This is demonstrated in the following case study.

The issue of informative censoring is not new and has been studied for decades. Yet it remains as one of the most frequent challenges in clinical data analysis today (DeMets, 2012). One of reasons for this remaining as a challenge is that determining whether censoring is related to patient's treatment response or not isn't always obvious. Lack of methods to account for different censoring mechanisms is another reason.

2. A Motivating Case Study

The challenge of informative censoring in time-to-success analysis is illustrated in the following case study. The case study is a phase III, randomized, double blind, parallel group study comparing a Test drug and an active control to assess non-inferiority. The efficacy endpoint is overall response at the end of treatment which is a binary composite endpoint with a value of either success or failure as assessed by an independent committee from multiple sources of data such as patients' clinical and mycological assessments. The primary efficacy analysis used a crude rate approach and results are displayed in Table 1.

Table 1 Treatment success rates at the end of treatment (EOT)

Overall response at EOT	*Test drug (N=199)*	*Control drug (N=201)*	*Adjusted Difference in Success Rate (95% CI)*
Success	61%	72%	-11% (-20%, -2%)
Failure	39%	28%	
Reasons of Failure			
Death	6%	5%	
Unsuccessful response	27%	21%	
Unevaluable	6%	2%	

Test drug had a significantly lower success rate than Control and noninferiority margin was not met – thus, noninferiority was not established. Upon further examination of the data, we noted that number of days on treatment until EOT varied greatly among patients with a range of 1 to 57 days for Test and 1 to 59 days for Control with a standard deviation of 12 days. Moreover, a high proportion of patients discontinued treatment without achieving success. It was therefore of interest to analyze the timing of success, rather than crude success rates, as earlier success has quality of life and economic benefits for patients. Initial analysis used the naïve Kaplan-Meier estimator, where success was treated as event and patients without observed success at EOT were censored at EOT. Figure 1 displays the plot of success rate (i.e., 1 - survival function) from the naïve KM analysis.

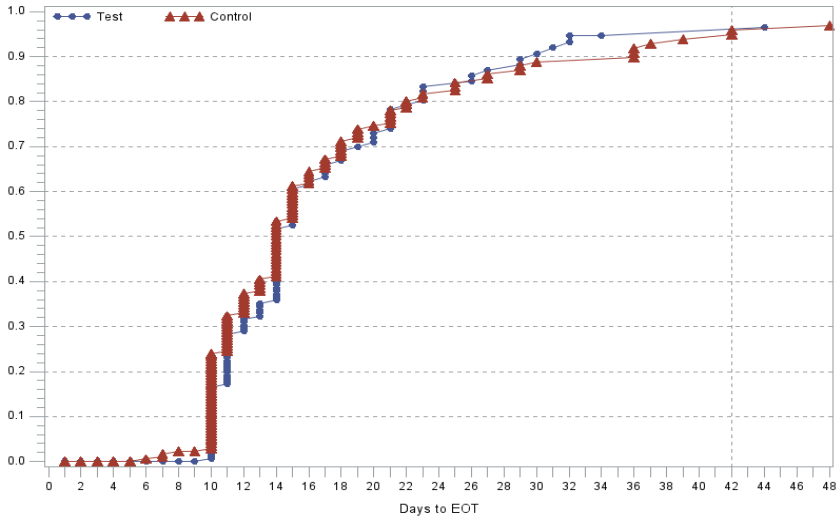


Figure 1 Naïve Kaplan-Meier plot for treatment success rate

The naïve survival analysis showed an estimated success rate of 95% for both drugs at reference time point (Day 42). Not only were these rates dramatically higher than the observed crude rates, but there was no difference in estimated success rate between the two groups which would have established noninferiority, a conclusion contradictory to the one from the crude rate analysis. The contradictory results indicate that the naïve KM estimator is biased for this data, most likely because many dropout patients did not meet the independent censoring assumption. In Section 3, we propose a two-step method to account for informative censoring.

3. Proposed Two-Step Method

3.1 Likelihood Index

Step one: Assign a parameter, which we call a likelihood index, denoted by *m*, for each censored subject. The value of *m* ranges from -1 to a positive number less than 1 which indicates a censored subject's likelihood of achieving success after censoring relative to similar subjects remaining in the study. It can also be considered as a measurement of the degree of departure from the assumption of independent censoring or missing at random (MAR). The value $m = 0$ corresponds to MAR or independent censoring which is the assumption of Kaplan-Meier estimator. The value $m = -1$ corresponds to zero likelihood or probability of achieving the event of success after censoring, while $0 < m < 1$ corresponds to subject having a higher likelihood of achieve success after censoring relative to otherwise similar subjects (i.e., similar with respect to observed outcome and baseline data) who are remaining in the study at the time of censoring. Each unique value of *m* represents a unique missing or censoring mechanism. The concept of likelihood index is analogous or similar to the shift parameter in pattern-mixture models (Daniels and Hogan, 2008) and the ignorability index (Ma *et al.*, 2005; Kaciroti *et al.*, 2012).

The process of modeling or assigning likelihood index *m* for censored subjects requires assumptions to be made which are usually untestable. As such, there is not a "correct" model to assign *m*. However, there are various plausible approaches.

In this paper, we demonstrate a non-modeling approach to assigning *m*. This approach is simple and practical enough for clinicians and non-statisticians alike to understand and apply. This method utilizes observed, often incomplete, outcome data by the time of dropout and reason for discontinuation to assign *m* for each censored patient. Table 2 illustrates the method. First, we start with the cases where the likelihood indices seem most obvious which we call anchor cases. In Table 2, Cases 1 and 7 are selected as anchors. We believe it is most reasonable to assign $m = -1$ for Case 1 which is a death on treatment. For Case 7 where the patient was alive and withdrew consent with missing both clinical and mycological assessments, it is reasonable to assign $m = 0$. Once the anchors are selected and assigned, we rank other cases relative to the two anchors with respect to their assumed likelihood of achieving success. For example, in Case 5 where patient had a clinical assessment of fail but a mycological assessment of success, it is reasonable to assume this patient has higher likelihood of success than Case 1 but lower likelihood than Case 7. Therefore, a value between -1 and 0 (we chose -0.25) is assigned to this case. In the last case where patient had successful mycological assessment but nonevaluable or missing clinical assessment, it is assumed that this patient has higher likelihood of success than Case 7. Therefore, a value larger than 0 (we chose 0.25) is assigned to this case.

Table 2 Assign likelihood index *m* for censored patients

Case #	*Death on EOT or EOT+1 day*	*Clinical assessment*	*Mycological assessment*	*Reason for discontinuation**	*Likelihood index (m)*
1	Yes	Any	Any	Any	-1
2	No	Fail	Fail or unevaluable	2	-1
3	No	Fail	Fail or unevaluable	1	-0.5
4	No	Unevaluable	Unevaluable	2	-0.5
5	No	Fail	Success	Any	-0.25
6	No	Success	Fail	Any	-0.25
7	No	Unevaluable	Unevaluable	1	0
8	No	Success	Unevaluable	Any	0.25
9	No	Unevaluable	Success	Any	0.25

*Reason for discontinuation: 1 for withdrew consent, 2 for otherwise

This assigning algorithm is based more on clinical evaluation than on statistical reasoning. It is therefore important to jointly develop with relevant clinicians and best do it in a blinded manner (i.e., without knowledge of subjects' treatment assignments), in order to minimize potential bias arising from the subjective element of the assigning model. Because there is no single correct model as assumptions about *m* are untestable, it is advisable to apply a number of different models that are based on reasonable assumptions as part of a sensitivity analysis.

3.2 Extended Kaplan-Meier (EKM) Estimator

Step two of the proposed method is to develop the extended Kaplan-Meier estimator to account for informative censoring utilizing the likelihood index *m*. Let us first review the Kaplan-Meier survival estimator. Let *N* denote the number of subjects at the beginning of the study and rank all subjects by the time of either event or censoring. The KM estimator of survival function can be formulated as follows:

$$\begin{aligned} &S_0 = 1, \\ &S_t = S_{t-1}\,(1 - I_t/R_t), \quad \text{for } t = 1,\ldots, N, \\ &R_{t+1} = R_t - 1, \qquad \text{for } I_t = 1 \text{ or } 0, \end{aligned} \tag{1}$$

where S_t is the survival function after the t-th subject experiences event or is censored, I_t is event indicator for the t-th subject (1 for event, 0 for censored), and R_t is the number of subjects remaining in the study just before the t-th subject experiences event or is censored (i.e., the "risk set").

If a subject t is censored under the mechanism of random censoring (or MAR), then $m_t = 0$ and the risk set will be reduced by 1. Let d_t denote the reduction of risk set following the censoring of subject t (i.e., $d_t = R_t - R_{t+1}$). In other words, if $m_t = 0$ then $d_t = 1$. If a subject t is censored under the assumption that the dropout has zero probability of an event (success) occurring at a later time, then $m_t = -1$ and it is appropriate to censor this subject at infinity time. Censoring a subject at infinity time means that the risk set is not affected by the censoring of this subject. In other words, if $m_t = -1$ then $d_t = 0$. Based on these two cases, we can generalize the formula for adjusting the risk set under all censoring mechanisms as follows:

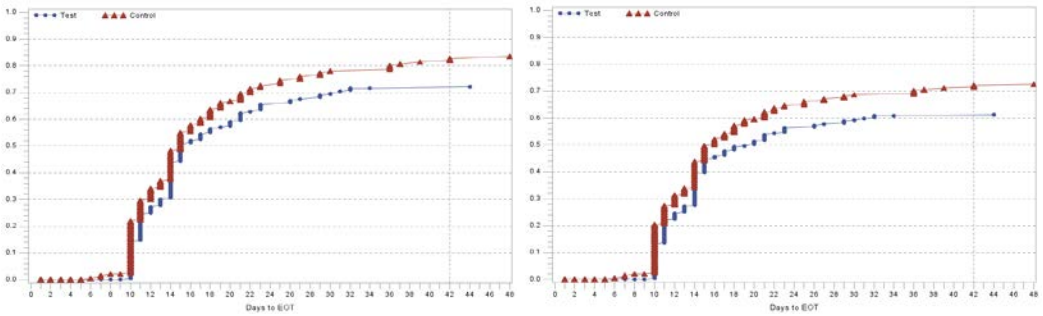
$$d_t = 1 + m_t. \tag{2}$$

Utilizing (2), we extended the KM estimator of survival function as follows:

$$\begin{aligned} &S_0 = 1, \\ &S_t = S_{t-1}\,(1 - I_t/R_t), \qquad \text{for } t = 1,\ldots, N, \\ &R_{t+1} = R_t - 1, \qquad \text{for } I_t = 1, \\ &R_{t+1} = R_t - d_t = R_t - (1 + m_t), \quad \text{for } I_t = 0. \end{aligned} \tag{3}$$

Note the difference between (1) and (3) which is the difference between standard KM estimator and our proposed EKM estimator. In the EKM estimator, when a subject t is censored, the risk set is no longer reduced by 1 but by $1 + m_t$. Under the KM assumption, $m = 0$ for all subjects, thus (3) becomes the same as (1) and the EKM estimator becomes the standard KM estimator. Therefore, the standard KM estimator is a special case of the EKM estimator. The advantage of the EKM estimator is that it allows for all possible values of m, thus it is capable of accounting for different missing mechanisms within the same study, provided that one can determine the underlying missing mechanisms for all subjects with reasonable accuracy and properly translate them into corresponding values of m. If a subject t has lower probability of experiencing event than a random dropout, in other words, $m_t < 0$, then $d_t < 1$. If another subject t+1 has higher probability of experiencing event than a random dropout, in other words, $m_{t+1} > 0$, then $d_{t+1} > 1$.

To illustrate the use of the EKM estimator, we apply it to the data from the case study in Section 2. Likelihood index m is derived for each censored subject using the algorithm in Table 2. Then EKM estimator is calculated using this set of m. This application was implemented using SAS version 9.3. The resulting cumulative incidence plot is displayed in Figure 2a. Here cumulative incidence function = 1 - S_t. Figure 2b displays the EKM curve under the most pessimistic assumption, that is, assuming all censored subjects will never have event ($m = -1$ for all censored subjects). It is easy to see that Figure 2b shows the same success rates as the initial crude rate analysis. The crude rate method is therefore another special case of the EKM estimator. By comparing Figures 2a and 2b, one can see that the extended estimator produced larger estimates of success rate for each treatment arm but difference between treatment arms is very similar with that from crude rate method. The EKM estimator thus supports the conclusion from the crude rate method.



a **b**

Figure 2 a) The proposed Extended Kaplan-Meier estimator, b) Under the assumption that $m = -1$ for all censored patients

Further sensitivity analysis can be performed using different sets of m derived from different models, instead of the model in Table 2.

4. Conclusions and Discussion

In this paper, we presented an extended Kaplan-Meier estimator that allows the flexibility to account for various censoring mechanisms for different patients, informative or noninformative, that can coexist in the same study, rather than making a single assumption for all censored subjects. Our proposed EKM estimator uses observed data from the dropouts to construct a unique likelihood index for each censored subject that quantitatively characterizes the underlying censoring mechanism or assumption for that subject, then utilizes the likelihood indices to adjust the risk sets. By contrast, the standard KM method ignores all recorded data of dropouts and simply assumes MAR for all dropouts. We showed that the EKM estimator contains both the standard KM estimator and the crude rate estimator as special cases. Assumptions on censoring mechanism cannot be verified by existing data. It comes down to which assumption(s) clinicians and statisticians believe are most reasonable and plausible. In the case study, it was clearly not reasonable to assume all dropouts were MAR. On the other hand, assuming all dropouts are "zero-chance" failure may not be reasonable either. Making a unique assumption for each subject based the subject's observed data is the most reasonable approach. This is the advantage of the proposed EKM estimator – it allows for various censoring mechanisms in the same study and for convenient implementation of sensitivity analysis using different sets of assumptions.

Compared to time-to-failure, time-to-success data is more prone to informative censoring by study design. While the proposed method was inspired by an analysis of time-to-success data, it is applicable to general time-to-event analysis where informative censoring is a concern. Because making assumptions on censoring mechanisms is subjective, algorithm for deriving m is best pre-specified without knowledge of subjects' treatment assignments. This can reduce potential bias.

Future relevant works include a more sophisticated model-based approach to assign m using not just observed outcome data but also baseline risk factors and other recorded data such as concomitant medications and reason for discontinuation of treatment. Confidence interval for the survival functions can be derived using adjusted risk sets in (3). An extended log-rank test may be similarly derived for treatment comparisons.

References

Daniels, M., Hogan, J. (2008). Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. Chapman & Hall/CRC.

DeMets, D.L. (2012). Current development in clinical trials: issues old and new. *Statistics in Medicine,* vol. 31, 2944-2954.

Kaciroti, N., Raghunathan, T.E., Taylor, J., Julius, S. (2012). A Bayesian model for time-to-event data with informative censoring. *Biostatistics,* vol. 13, 341-354.

Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, vol. 53, 457-481.

Ma, G., Troxel, A.B., Heitjan, D. (2005). An index of local sensitivity to nonignorable drop-out in longitudinal modeling. *Statistics in Medicine,* vol. 24, 2129–2150.