Nonparametric Method for Analyzing Recurrence of Adverse Events in Randomized Clinical Trials

Masanori Ito[1], Misun Lee[1]
[1]Astellas Pharma Global Development Inc., 1 Astellas Way, Northbrook, IL 60062

Abstract

Adverse events (AEs) from randomized clinical trials are often summarized as a crude incidence rate based on the percentage of patients who had at least one occurrence of AE or by an exposure-adjusted incidence rates. The AEs that are reported multiple times may be inadequately addressed by the crude summary. An approach with mean cumulative function (MCF) can assess multiple or reoccurring AEs to subjects. A method was proposed by taking the MCF of AEs which is a similar approach to a Cox model. Such approaches can assess the AE profiles between the treatment groups, however, the AE occurring days from either treatment groups are considered for both treatment groups based on the regression model. This may misrepresent the actual occurrence days of AEs. Furthermore, the assumption of proportionality between treatment groups is not often valid. Therefore, we propose a nonparametric MCF method as an additional approach in conjunction with the widely used crude incidence approach. We investigated the results from a randomized Phase 3 study and the results are presented. Throughout simulation, we further investigated s and graphical approaches are presented to compare the semi-parametric MCF method with our proposed nonparametric MCF method.

Key Words: Recurrent Adverse Events, Mean Cumulative Function, Kaplan-Meier, Non-parametric, Modified Log-rank Test

1. Introduction

As part of safety evaluations of an investigational product from randomized clinical trials, the adverse events (AEs) data are collected and summarized. The most widely used method evaluating the AEs is crude incidence rates or exposure adjusted incidence rates. The patients who reported AEs are counted once for a given term such as System Organ Class (SOC), High Level Terms (HLT) or Preferred Terms (PT) even if patients may have reported multiple events within the term or reported the event multiple times. With such an approach, difference in crude rates between treatment groups could be biased when: a) discontinuation rates are different between treatment groups, or b) recurrent AEs are not considered. It is wrong to ignore when multiple AEs are reported by a patient or same AE is reported multiple times by a patient. As an alternative to the crude incidence rate or exposure adjusted incidence rate, a mean cumulative function (MCF) approach is used in order to address the recurrent AEs during clinical trials. Lawless and Nadeau (1995) explored MCF and discussed details on how to estimate the MCF and its robust variance. Siddiqui (2009) evaluated AEs from phase 3 trial in various ways and compared the results. Cao and He (2011) applied MCF to the AEs data based on the semi-parametric model similar to the proportional hazard model and compared MCFs between treatment groups. A semi-parametric MCF approach can compare the overall

recurrent AEs rates between the treatment groups based on a Cox model. However, some drawbacks from the semi-parametric MCF approach are observed. In this paper, we discussed those issues. We applied different approach to the actual AE data from a Phase 3 randomized study and results are presented in Section 3. We compared the test results from the Fisher's exact test, semi-parametric MCF and non-parametric MCF. Our simulation results that investigated further between the semi-parametric MCF and nonparametric MCF are presented in Section 4.

2. Mean Cumulative Function Approaches

2.1 Semi-parametric MCF Approach

As an alternative to the crude incidence summary, a mean cumulative function approach can be used to analyze multiple or repeated AEs to patients. The recurrent AE data consist of the same event or different events occurring at different times for each patient. Examples of AE profiles for patients and MCF curve for all patients in a trial are depicted in Figure 1. The staircase functions that are indicated by the dotted lines are individual patients cumulative AEs profiles. A simplistic way to visualize an MCF for a study is depicted by a darker dotted line which is an average curve from all subjects' cumulative AE profiles.

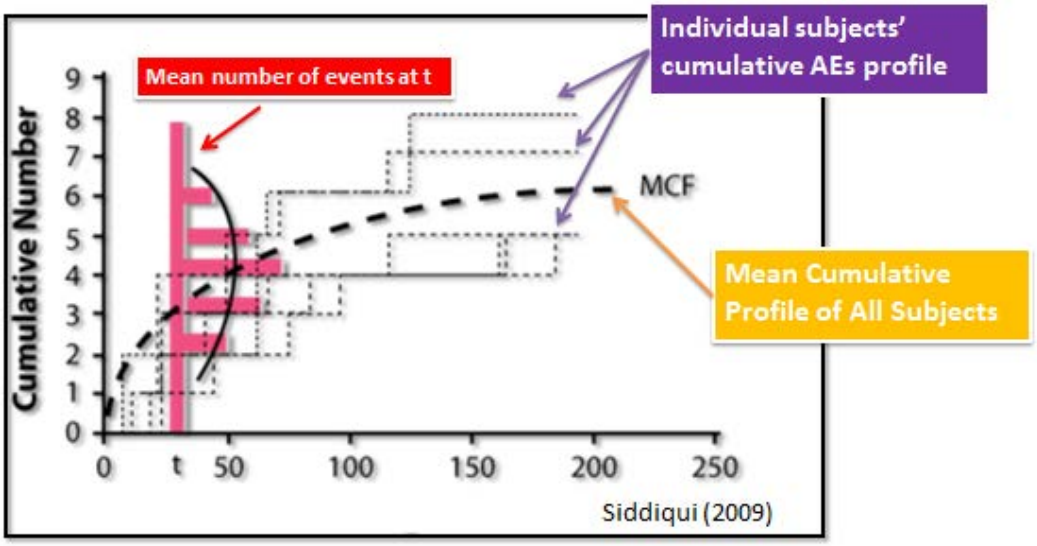

Figure 1: Plot of AE Profiles and MCF

Cao and He (2011) applied semi-parametric MCF to the AEs data and compared the MCFs between treatment groups.

Assume *M(t)* is a mean cumulative number of AEs up to time *t*. The estimate of MCF at time *t* can be expressed as

$$\widehat{M}(t) = M_0(t)exp(X'\beta), \qquad (1)$$

where, $M_0(t)$ is a baseline MCF at time t, X′ is a time invariant covariates vector, and β is a slope for treatment effect. The test for the null hypothesis $H_0: \beta = 0$ can be conducted against the alternative hypothesis $H_1: \beta \neq 0$. The semi-parametric MCH approach may compare the treatment effect fairly easily, however, the AE occurrence

days from either treatment groups are considered for both treatment groups based on the proportional hazard model. This may misrepresent the actual occurrence days of AEs from either treatment groups. As an example, **Figure 2** shows the MCF estimates for two treatment groups from a semi-parametric model.

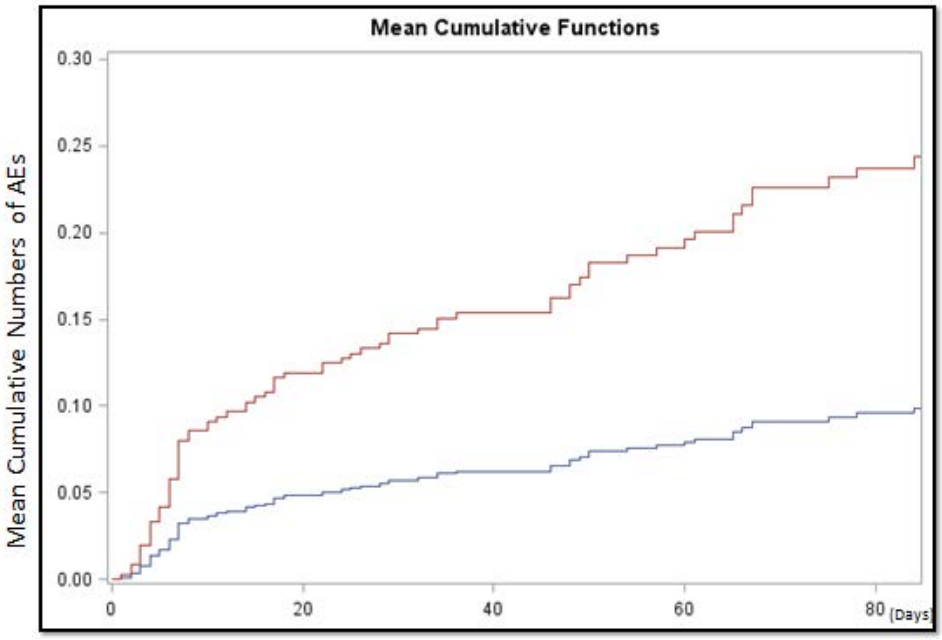

Figure 2: Example of MCF estimates from a semi-parametric model

These two plots suggest a drawback of this approach: they have exact same pattern of staircase jump. Additionally, the assumption of the proportionality between treatment groups is not often valid for the AEs data in the clinical trial.

2.2 Non-parametric Mean Cumulative Function Approach

Let $n_{ji}(s) \geq 0$ be the number of recurrences for subject i within time $[0, t]$, where $j =$ treatment group. Then $m_j(t) = E[n_{ji}(t)]$ and $M_j(t) = \sum_{s=0}^{t} m_j(s)$. Subject ji is on treatment or followed over $[0, \tau_{ji}]$, and for notational convenience, we define $\delta_{ji}(t) = I(t \leq \tau_{ji})$.

The total number of recurrences and total number of subjects observed by time t, respectively, are denoted by $n_{j\cdot}(t) = \sum_{i=1}^{I} \delta_{ji}(t) n_{ji}(t)$ and $\delta_{j\cdot}(t) = \sum_{i=1}^{I} \delta_{ji}(t)$. If the $n_{ji}(t)$ (t =0, 1, 2,…) are independent Poisson random variables with means $m_j(t)$, the maximum likelihood estimate of $m_j(t)$ is $\hat{m}_j(t) = n_{j\cdot}(t)/\delta_{j\cdot}(t)$, and the estimate of $M_j(t)$ for $0 \leq t \leq \tau$ is

$$\hat{M}_j(t) = \sum_{s=0}^{t} \hat{m}_j(s) = \sum_{s=0}^{t} \frac{n_{j\cdot}(t)}{\delta_{j\cdot}(t)} \qquad (2)$$

Equation (2) is referred to as the Nelson-Aalen estimator (Andersen and Borgan, 1985) and it is well known as a non-parametric estimate in counting process. In comparing two treatment groups ($j = 0, 1$), the score statistics on the null hypothesis $H_0: M_0(t) = M_{01}(t)\ for\ all\ 0 \leq t \leq \tau$ is

$$U = \sum_{s=0}^{\tau} \frac{\delta_0.(s)\delta_1.(s)}{\delta_0.(s) + \delta_1.(s)} \left\{ \frac{n_1.(s)}{\delta_1.(s)} - \frac{n_0.(s)}{\delta_0.(s)} \right\} \quad (3)$$

The first term in the summation in equation (3) can be considered as the weight based on the number of at risks at time *s* from two treatment groups. It is analogous to a log-rank statistic for testing the equality of survival distributions. A robust variance for a score statistic of (3) can be calculated by the following formula

$$\widehat{Var}(U) = \sum_{j=0}^{1} \sum_{i=1}^{I_j} \left\{ \sum_{s=0}^{\tau} \delta_{ij}(s) \frac{\delta_{..}(s) - \delta_{j.}(s)}{\delta_{..}(s)} \left[n_{ji}(s) - \frac{n_{j.}(s)}{\delta_{j.}(s)} \right] \right\}^2 \quad (4)$$

where $n_{ji}(s)$ is the number of recurrences at time s for the $i'th$ subject of treatment j ($j = 0,1$), $\delta_{ji}(s)$ indicates whether the subject is observed at time s, and dots indicate summation over the appropriate indices. We can use $U^2/\widehat{Var}(U)$ as the test statistic which has a Chi-square distribution with a 1 degree of freedom.

3. Case Study

Using the actual AE data from a randomized Phase 3 clinical trial, we compared the results from the various methods. This study was conducted for patients who were immunocompromised thus high AE rates were expected. We use dummy treatment labels Trt A and Trt B throughout this paper. The summary and statistical testing for treatment group comparisons were performed on the summary for the SOC term (**Table 1**). We compare the results from three approaches: a Fisher's exact method for crude rates, semi-parametric MCF and non-parametric MCF. The summary table also includes crude incidence number and percentage by counting patients within the relevant term once, 95% CI for each treatment group, number of incidence which counts all events within the relevant term, relative risk, and p-values for the treatment groups comparison from each method.

Firstly, we assessed the treatment discontinuation rates and the treatment duration to ensure for no apparent bias from treatment discontinuation or duration. The figures suggest that the treatment groups were balanced and no suggestion of bias from this aspect. . In general, the difference of discontinuation rate and exposure duration between treatment groups could be biased for the AEs analysis. **Figure 3** shows the time-to-discontinuation rates the histogram of treatment durations for two treatment groups. These plots indicate good balance between the treatment groups for discontinuation rates and exposure durations.

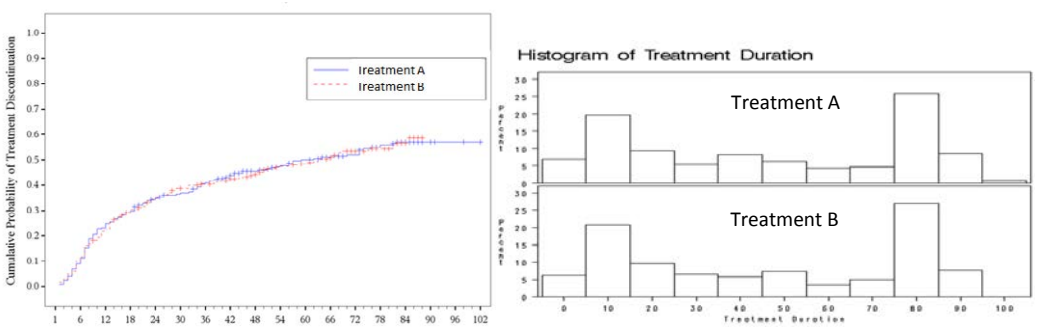

Figure 3: Cumulative rate of treatment discontinuation and histogram of treatment duration for two treatment groups

Table 1: Adverse Events Crude rates, Relative Risks, and P-values by Various Methods from a Randomized Clinical Trial

Overall/SOCs	Trt A (%) [95% CI] <N of AEs> (N=257)	Trt B (%) [95% CI] <N of AEs> (N=259)	Relative Risk (Trt A/Trt B) and 95% CI	P-values		
				Fisher's exact	SMCF	NMCF
Overall	247 (96.1%) [93.0%, 98.1%] <2829>	255 (98.5%) [96.1%, 99.6%] <3463>	0.98 [0.95, 1.01]	0.112	0.007*	0.122
Hepatobiliary disorders	23 (8.9%) [5.8%, 13.1%] <23>	42 (16.2%) [11.9%, 21.3%] <61>	0.55 [0.34, 0.89]	0.016*	<0.001*	<0.001*
Cardiac disorders	43 (16.7%) [12.4%, 21.9%] <58>	57 (22.0%) [17.1%, 27.6%] <76>	0.76 [0.53, 1.09]	0.148	0.159	0.014*
Gastrointestinal disorders	174 (67.7%) [61.6%, 73.4%] <487>	180 (69.5%) [63.5%, 75.1%] <618>	0.97 [0.87, 1.10]	0.705	0.038*	0.041*
Skin and subcutaneous tissue disorders	86 (33.5%) [27.7%, 39.6%] <129>	110 (42.5%) [36.4%, 48.7%] <160>	0.79 [0.63, 0.99]	0.037*	0.100	0.299

*: two sided p-value $\leq$0.05. Numbers in the brackets are the counts of incidences.
SMCF: Semi-parametric MCF, NMCF: Non-parametric MCF

The overall incidence rates and selected SOC terms are included in Table 1. We explore details of the distribution for each situation below:

- For Hepatobiliary Disorders, all approaches identified statistical difference between the treatment groups with a p-value<0.05. This finding is also supported by the relative risk 0.55 and its 95% CI (0.34-0.89).

- For Cardiac Disorders, both Fisher's test and semi-parametric MCF provided non-significant p-values although the nonparametric MCF provided the significant p-value. To further investigate, data distribution is shown in **Figure 4**.

Case: P-value<0.05 from only Non-parametric MCF test – the crude incidences of n=43 and n=57 patients were not seen statistically different however the plots below suggest that the Treatment B has more AEs reported at a later time of the trial. Without the non-parametric MCF approach, this assessment could have been overlooked.

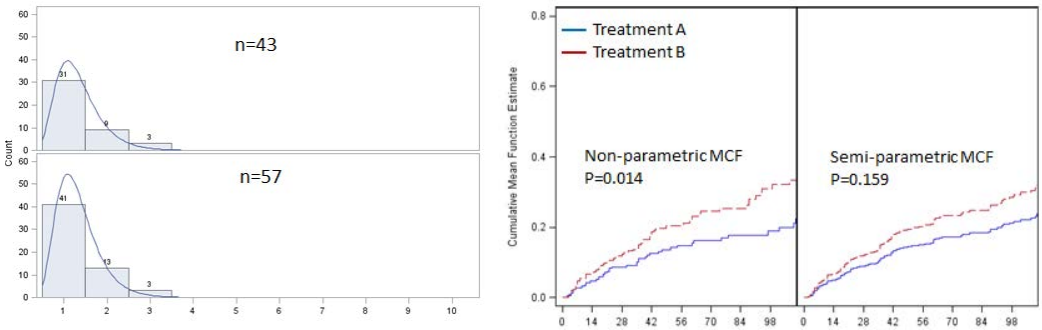

Figure 4: Distributions for number of events and cumulative plot from MCFs (Cardiac disorders)

- For Gastrointestinal Disorders, the widely used Fisher's exact test provided non-significant p-value while both MCF methods provided significant p-values. Further investigation is done through the data distribution shown in **Figure 5**.

Case: P-value>0.05 with Fisher's test however p-value<0.05 from MCF tests

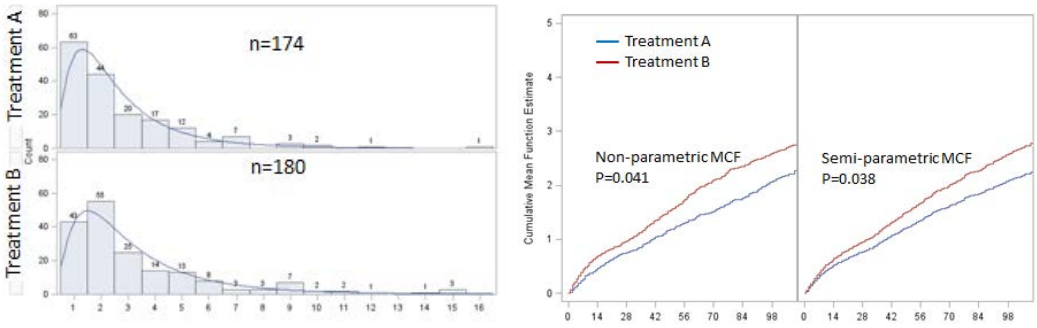

Figure 5: Distributions for number of events and cumulative plot from MCFs: (Gastrointestinal disorders)

- For Skin and Subcutaneous Tissue Disorders, only the Fisher's test provided significant p-value and both two MCFs did not provide significant test result. Data distribution is shown in **Figure 6**.

Case: P-value<0.05 with a Fisher's test and p-value>0.05 from MCF tests

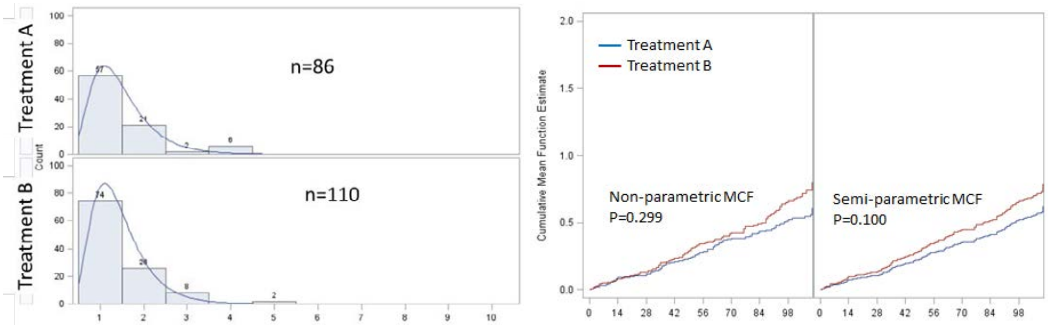

Figure 6: Distributions for number of events and cumulative plot from MCFs (Skin and subcutaneous tissue disorders)

4. Simulations

From the case study results presented in the previous section, we experienced a situation where the non-parametric MCF method provided significant treatment difference although the semi-parametric MCF method did not so. In order to further explore whether the sensitivity of the MCF approaches would depend on the proportionality of the data, we performed simulations. In the simulation setting, AEs were assumed to occur based on a piecewise Weibull distribution with the following form,

$$\boldsymbol{f_{jk}(t)} = \frac{\boldsymbol{a_{jk}}}{\boldsymbol{b_{jk}^{a_{jk}}}} \boldsymbol{e}^{-\left(\frac{t}{b_{jk}}\right)^{a_{jk}}}, \qquad (5)$$

where, t is the time from previous event occurred, $j = 0,1$ is the treatment code (test drug or comparator), and $k = 1,2$ is time period (early period or later period in the clinical trial). a_{jk} and b_{jk} are the shape and scale parameters, respectively.
Four scenarios are considered in this simulation by the combination of the treatment group and the time period:

a) No treatment difference in AEs (to check the Type I error rate)
b) Difference in AEs with proportionality of cumulative function between treatment groups
c) Difference in AEs without proportionality of cumulative functions: AEs occurred frequently in early period in the comparator group
d) Difference in AEs without proportionality of cumulative functions: AEs occurred frequently in later period in the comparator group

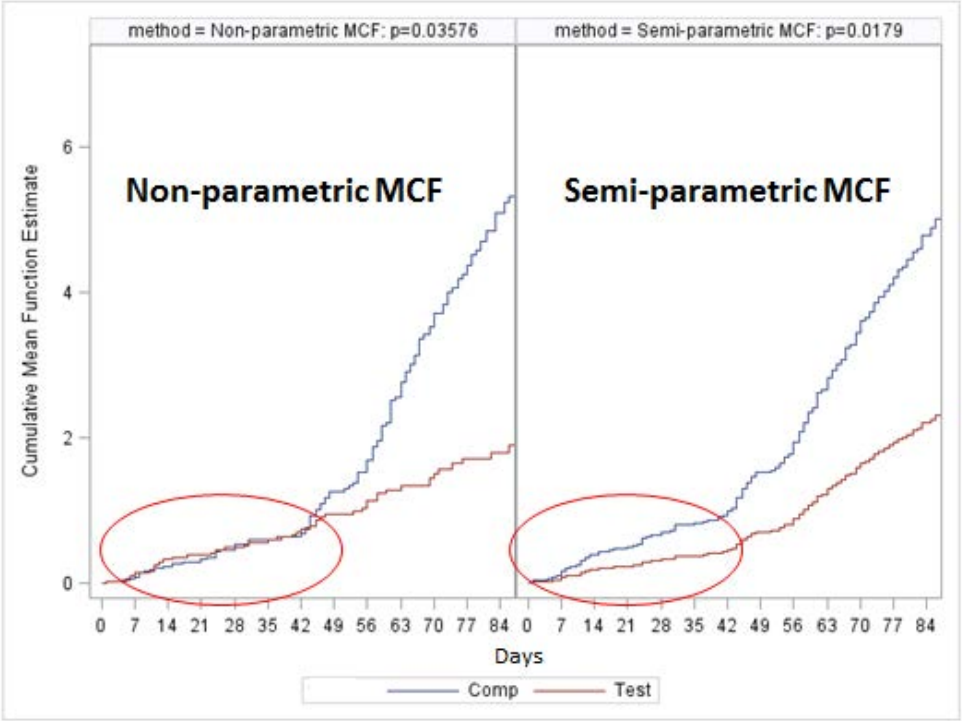

Figure 7: Simulation example of cumulative plots from MCFs in scenario d)

For these scenarios, the censoring was generated from a uniform distribution. Sample size was set at 50/arm, and the semi-parametric and non-parametric MCF approaches were applied. Statistical power of two MCFs was calculated and compared from 500 times simulation. **Table 2** shows the summary of the simulation results from four scenarios. The powers from both MCFs approaches were smaller than 5% from the Scenario A, which ensures no inflation of a Type I error rate. In the Scenario B, the powers from the MCF approaches were similar with 83-84%. This is as expected as the AE proportionality between treatment groups was present. From Scenario C, The power from the non-parametric MCF method was higher than that from the semi-parametric MCF method which AEs occurred frequently in earlier period in the comparator group. It suggests that the non-parametric MCF method was more sensitive than the semi-parametric MCF method in detecting the difference. On the other hand, the semi-parametric MCF method was better than non-parametric in Scenario D. **Figure 7** shows MCF estimates from one of the simulation result from Scenario D. The estimates from the non-parametric MCF show no difference in the early part which is consistent to the actual data however the estimates from the semi-parametric MCF method suggests treatment difference even during the early time period that is influenced by the treatment difference in the later time period.

Table 2: Summary of the simulation results

Scenario	Treatment	Time interval *t*	Scale parameter	Power of Semi-parametric MCF	Power of Non-parametric MCF
Scenario a)	Test drug	[Day 1, Day 84]	2	4.4%	4.2%
	Comparator	[Day 1, Day 84]	2		
Scenario b)	Test drug	[Day 1, Day 84]	2	83.6%	84.8%
	Comparator	[Day 1, Day 84]	0.8		
Scenario c)	Test drug	[Day 1, Day 84]	2	75.6%	80.8%
	Comparator	[Day 1, Day 24]	0.7		
		[Day 24, Day 84]	2		
Scenario d)	Test drug	[Day 1, Day 56]	2	66.6%	47.8%
	Comparator	[Day 1, Day 56]	2		
		[Day 56, Day 84]	0.1		

* Shape parameter was 1 for all scenarios in the Weibull distribution.

5. Discussion and Conclusions

In this paper, we examined strengths and weaknesses of the semi-parametric and non-parametric MCF approaches using the Phase 3 data and simulations. We presented the cases where the crude incidence rates and its treatment groups are found to be not significantly different from the Fisher's exact test however, the recurrence of AE may be significantly different between the treatment group and this difference is detected by the MCF method. Therefore, the statistics such as P-values and confidence intervals may not adequately explain safety profile of AEs. Based on these results, we concluded that we should not limit to one or two approaches when evaluating AEs. In the MCF approaches, we need to have caution when presenting AE profiles by graph from semi-parametric approach as we discussed in section 2.1. In this work, we don't concern about possible multiplicity issue as the primary focus of the AE analysis is to examine the safety of drug thoroughly and not to overlook the AE recurrence.

We strongly recommend using both non-parametric MCF and semi-parametric MCF methods as sensitivity analysis. We also challenge the widely used method of analyzing AEs by crude incidence rate alone when statistical inference is made. In order to thoroughly evaluate the AEs from clinical trials, we should not limit to one or two methods and use different types of available statistical methods. We should also plot the data to understand the real pattern of AEs. An example is shown in the Figure 8. The crude incidence rates are equal (n=15) from both treatment groups however, their AE profiles are drastically different.

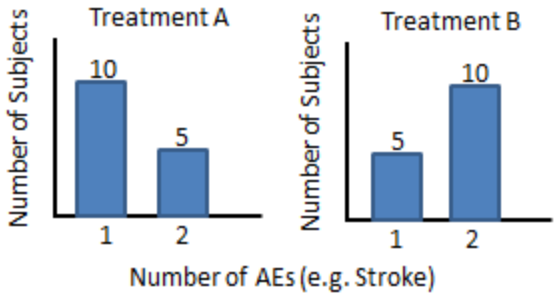

Figure 8: Example of events where crude incidence is same while complete profile is drastically different

In order to thoroughly examine the AE data, we strongly recommend using the MCF approach in conjunction with the crude incidence rates. Number of incidences can be also informative although it has its limitation. Since the MCF approach is a stochastic counting process of a cumulative number of events, it may be an acceptable statistical approach for understanding the complete AE profiles of treatments from randomized clinical trials.

References

Andersen, P. K. and Borgan, O. (1985). Counting Process Models for Life History Data: A Review. Scandinavian Journal of Statistics, 12, 97-158.

Lawless, J. F. and Nadeau, C. (1995). Some Simple Robust Methods for the Analysis of Recurrent Events. Technometrics, 37, 158-168.

Yamaguchi, T. (2005). Recurrent Events Data Analysis: A Review. Japanese Journal of Biometrics, 26 (2), 81-117.

Siddiqui, O. (2009). Statistical Methods to Analyze Adverse Events Data of Randomized Clinical Trials. Journal of Biopharmaceutical Statistics, 19, 889-899.

Cao, D. and He, X. (2011). Statistical Analysis of Adverse Events in Randomized Clinical Trials using SAS. PharmaSUG2011, SP07.