Statistical Consideration for Event-Free Survival With Cure Rate in Acute Myeloid Leukemia Studies

Yuichiro Kaneko[1] (yuichiro.kaneko@astellas.com) , Kentaro Takeda[2], Shufang Liu[3] , Lu Tian[4]
(1: Astellas Pharma Inc., 2: Astellas Pharma Global Development Inc. 3: Gilead Sciences, 4: Stanford University)

Introduction

Acute myeloid leukemia (AML) is a life-threatening, aggressive malignancy of the white blood cells, and the most common form of acute leukemia in adults.
The clinical studies for newly diagnosed AM patients often starts from induction phase through the consolidation phase, or even through the maintenance phase. Through the recent rapid progress of AML treatment with achievement of high CR rate, event-free survival (EFS) played an important role in AML research and has been accepted as a primary efficacy endpoint in some AML studies.
With the issue of the FDA guidance on AML, EFS is clearly defined as the time from the date of randomization to the date of:

- Induction treatment failure (ITF),
- Relapse for those who have induction treatment success (for example, complete remission (CR)), or
- Death from any cause,

whichever comes first.

From the statistical analysis viewpoint for EFS, the proportional hazard assumptions no longer applies on the whole, because ITF events account for a certain fraction of the total events in general.
Recent research by Xu et al. (2021) provides the statistical methodology for decomposing the patient population into those with ITF events and those contributing relapse or death events after the randomization date by partitioning the conventional log-rank test statistic as a linear combination of statistics for those two portions.

Motivation Example

As seen especially in the newly diagnostic AML studies with curative intent shown in midostaurin phase 3 study (RATIFY), there was some indication for the cure fraction in the patients who achieved complete remission and **cured** thereafter, which was not being taken into consideration in the research of Xu et al. (2021).

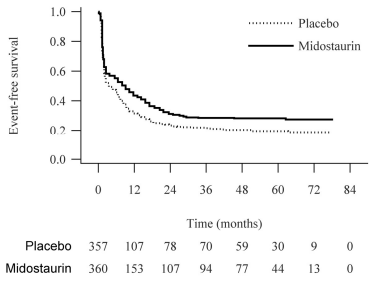


FIGURE 1 Kaplan-Meier Curve for Event-Free Survival in RATIFY study (reconstruction)

Objective

Our research aims to further investigate the relationship between sample size and statistical power, considering the cure rate and examining the operational characteristics by simulations focusing on studies with newly diagnosed AML patients.

Partition of the log-rank test by Xu et al. (2021)

Xu et al. (2021) partitioned log-rank test into ITF and Non-ITF components. The expected value of ITF portion is expressed as follows.

$$E(Z_{D1}) = \frac{\sqrt{\theta(1-\theta)n_{tot}}}{\sqrt{P_{CR}(1-P_{CR})}}\left(P_{2,CR} - P_{1,CR}\right) \quad \ldots\ldots(1)$$

$P_{1,CR}$, $P_{2,CR}$, P_{CR} denote CR rate in the treatment group and control group, and overall. Also, θ denote as allocated proportion in the treatment groups.

Our proposed approach

Our proposed approach followed the partition of log-rank test for ITF and Non-ITF, and further partition Non-ITF (CR responders) into 'cure' or 'non-cure' part.

Three Components for Event-Free Survival

1. ITF portion: No CR
   - The patients who failed to achieve CR.
2. Cured portion:
   - The patients who achieved CR with no possibility of relapse, death.
3. Patients who achieve CR but not cured portion:
   - The patients who achieved CR but not in the cured portion.

Mathematical Formulation for proposed approach

Our proposed approach partitioned into log-rank test in the same way for ITF (Component 1) as shown Equation (1). Then, for Non-ITF components, we proposed to divide further by incorporating the cure rate model.
Note: $S^*_{-D1,i}(t) = \pi_i + (1-\pi_i)S_{-D1,i}(t)$,

$$\lambda^*_{-D1,i}(t) = \frac{(1-\pi_i)S_{-D1,i}(t)}{\pi_i+(1-\pi_i)S_{-D1,i}(t)}\lambda_{-D1,i}(t)$$

G_{-D1}: common censoring distribution for Non−TF part
n_{21}, n_{22} is the number of patients after Day 1 in treatment, control group, respectively, and $n_2 = n_{21} + n_{22}$ and $\theta' = \lim_{n_2\to\infty} \frac{n_{21}}{n_2}$,
π_i: cure rate in ith group, τ: study duration (month)
$\lambda_{-D1,i}(t)$: survival function for Non-TF part non-cured patients in ith group
Then, the test statistics for Non-ITF components (Component 2, 3) are derived in the following equation.

$$E(Z_{-D1}) = \frac{\mu_{-D1}}{\sigma_{-D1}/\sqrt{n_2}} \quad \ldots\ldots(2)$$

where μ_{-D1}, σ^2_{D1} are derived as

$$\mu_{-D1} = \theta'(1-\theta')\int_1^\tau \frac{S^*_{-D1,1}(t)\cdot S^*_{-D1,2}(t)}{S^*_{-D1}(t)}\cdot G_{-D1,1}(t)\left\{\lambda^*_{-D1,2}(t) - \lambda^*_{-D1,1}(t)\right\}dt$$

$$\sigma^2_{-D} = \theta'(1-\theta')\int_1^\tau \frac{S^*_{-D1,1}(t)\cdot S^*_{-D1,2}(t)}{\left\{\theta' S^*_{-D1,1}(t) + (1-\theta')S^*_{-D1,2}(t)\right\}^2} G_{-D1,1}(t)$$
$$\left\{\theta' S^*_{-D1,1}(t)\lambda^*_{-D1,1}(t) + (1-\theta')S^*_{-D1,2}(t)\lambda^*_{-D\ ,2}(t)\right\}dt$$

Then, from the Equation (1) and (2), the corresponding log-rank statistics is expressed as

$$Z^*_{lr} = \sqrt{W^*}Z_{D1} + \sqrt{1-W^*}Z^*_{-D}\ .$$
$$\Phi\left\{E(Z^*_{lr}) - z_{\alpha/2}\right\} = 1 - z_\beta,$$

where the corresponding weight is expressed

$$w^* = \frac{d_1(n_{tot}-d_1)}{n_{tot}}\theta(1-\theta)\Big/\left\{\frac{d_1(n_{tot}-d_1)}{n_{tot}}\theta(1-\theta) + \theta'(1-\theta')\right.$$
$$\left.\int_1^\tau \frac{S^*_{-D1,1}(t)\cdot S^*_{-D\ ,2}(t)}{\left\{\theta' S^*_{-D1,1}(t) + (1-\theta')S^*_{-D1,2}(t)\right\}^2} G_{-D1,1}(t)\right.$$
$$\left.\cdot\left(\theta' S^*_{-D1,1}(t)\lambda^*_{-D1,1}(t) + (1-\theta')S^*_{-D1,2}(t)\lambda^*_{-D1,2}(t)\right)dt\right\}$$

Simulation in Motivation Example

To assess the impact of the cure fraction on the statistical power and the accuracy of proposed estimation of statistical power of log-rank test in the presence of non-zero cure portion, we mimicked the motivation example with the same number of events (D=540) but expanded to incorporate positive cure rate.

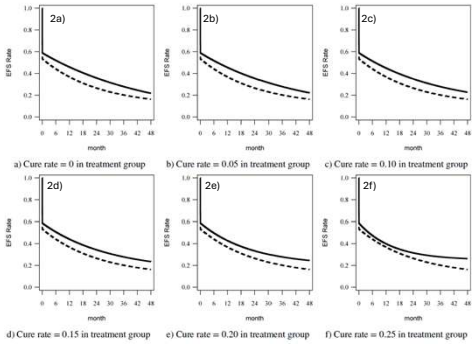


Figure 2. EFS Kaplan-Meier curve with the different patterns for cure rate being (0.10, 0), (0.10, 0.05), (0.10, 0.10), (0.10, 0.15), (0.10, 0.20), (0.10, 0.25) in the control and treatment groups, while 4-year EFS rate was fixed at 0.206 and 0.282 for the control and the treatment group

In Figure 3, the statistical power from our proposed analytical calculation is similar to the one from simulation based on the conventional log-rank test under different cure rate in the treatment group. In addition, the statistical power decreased from 0.755 in Figure 2a to 0.519 in Figure 2f with an increase in the cure rate in the treatment group even under the fixed number of events.

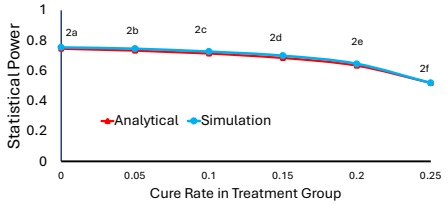


Figure 3. Analytically calculated power based on Z^*_{lr} and simulated statistical power based on the conventional log-rank test

New Simulation Study Design

We simulated the trial 100,000 times with an equal allocation of patients to a treatment group or a control group with the total sample size ranged from 100 to 1000 patients by altering each parameters from base case EFS function in Figure 4 to assess the impact of these parameters.

- The study duration was set to be 48 months with a 24-month accrual period and a 24-month follow-up period.
- The CR was defined as the achievement of CR criteria by 24 weeks.
- The CR rate was set as (60%, 50%) in the treatment and control group.
- The cure rate was assumed t treatment and control group.
- HR is set as 0.8.

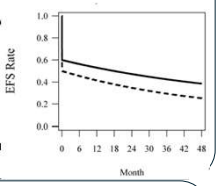


Figure 4. Base Case EFS function with fixed CR rate = (0.5, 0.6) and fixed cure rate = (0.1, 0.2), in control group vs. treatment group, and HR=0.8 for Patien who achieve CR but not cured portion

Simulation Result (1): Assessment of Analytical Approach vs. Simulated Approach

The analytically calculated power estimated by our proposed method was similar to the statistical power in simulation based on the conventional log-rank test. Moreover, the statistical power for both Z_{D1} and Z^*_{-D1} increased as the total number of patients increased, while the weight for the test statistic was almost constant (ranged from 0.568 to 0.570) across the different number of patients, and this weight was almost identical to the estimated weight by the proposed approach which is constant with 0.572.

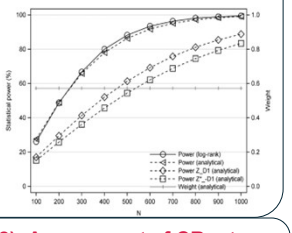


Figure 5. Statistical power and total number of patients in simulation study. The analytical weight value is displayed for right vertical axis.

Simulation Result (2): Assessment of CR rate

The statistical power was highly increased as the number of patients increased and the degree of CR rate difference between groups increased.

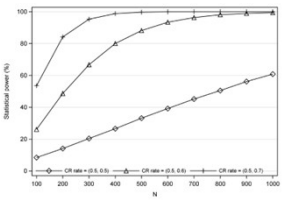


FIGURE 6 Statistical power based on the conventional log-rank test and total number of patients in the simulation study when the CR rate was varied from (0.5, 0.6) as default to (0.5, 0.5), (0.5,0.7)

Simulation Result (3): Assessment of cure rate

The statistical power was also highly impacted by the change in cure rate. When the cure rate in the treatment group was higher than that in the control group, the statistical power was higher.

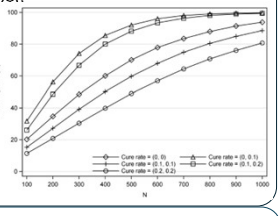


FIGURE 7 Statistical power based on the conventional log-rank test and total number of patients in the simulation study when the cure rate was varied from (0.1, 0.2) as default to (0, 0), (0.1, 0.1), (0.2, 0.2)

Conclusion

- The statistical power analytically calculated by our proposed method was similar to the statistical power from simulation based on the conventional log-rank test. This means the statistical power for EFS with 3 components can be analytically computed without simulation.
- Our proposed new estimator for the power can be an useful analytic tool to investigate the dependence of the study power on parameters such as CR rates, cure rates, and hazard ratio among patients who achieve CR but not cured in the stage of study design.

Reference

Xu T, Zhang Y, Goldwasser M, Zhang V. Statistical considerations in clinical trial design with event-free survival as the primary efficacy endpoint. *Pharmaceutical Statistics* 2021; 20(1): 721–736.