

Feasibility of New Method for the Prediction of Clinical Trial Results Using Compressive Sensing of Artificial Intelligence - GOTIC Study

Yasunari Miyagi

Representative, Medical Data Labo, 289-48 Yamasaki,
Naka Ward, Okayama City, Okayama Prefecture 703-8267, Japan
Director, Department of Gynecology, Miyake Ofuku Clinic, 393-1 Ofuku,
Minami ward, Okayama City, Okayama Prefecture 701-0204, Japan

Keiichi Fujiwara

Department of Gynecologic Oncology, Saitama Medical University
International Medical Center, Hidaka city, Saitama prefecture, Japan,
Department of Obstetrics and Gynecology, International University of
Health and Welfare, Narita city, Chiba prefecture, Japan

Hisanaga Nomura

Department of Data Science / Pharmacy, National Cancer Center Hospital East,
Kashiwa City, Chiba prefecture, Japan

Koji Yamamoto

Department of Biostatistics, Yokohama City University School of Medicine,
Yokohama city, Kanagawa prefecture, Japan

Robert L Coleman

US Oncology Research, The Woodlands, TX, USA

Abstract

To develop a new method for interim analysis to obtain the p-values distribution profile of the log-rank test and the statistical power at various information fraction by iteratively predicting the lacking time data of censored and uncensored cases respectively for each arm at the time when the determined sample size will be achieved. A compressive sensing algorithm derived from artificial intelligence was developed for comparing two groups of clinical trial data from real-world databases of deidentified individual participant-level data. The judgement of the interim analysis was compared to the conventional method, with our method demonstrating that it was at least comparable to the conventional method. The power by our method was higher than the conditional power of the conventional method. When α -error by log-rank test was focused, our method seemed to have made the right decision earlier than the conventional method. It may be possible to expand the options of applicable methods for interim analysis. It would be helpful in establishing efficacy and futility judgement in clinical trials.

Keywords: Clinical Trial; Artificial Intelligence; Compressive Sensing; Computer Science; Study Design.

INTRODUCTION

In clinical trials of cancer, progression-free survival (PFS) and overall survival (OS) are important primary endpoints. The changes in events over time are analyzed as important endpoints and commonly displayed using Kaplan-Meier curves. In particular, the log-rank test is often used as the basic test for significant differences in the survival curves between two groups. The optimal two-stage design was proposed by Simon. This method determines the adequate sample size by calculating the response probability of two groups with α and β errors [1,2]. Cancer treatment decisions and drug development have traditionally relied on statistical methods using these errors parameters.

While these methods may help accelerate hypothesis testing, successful clinical trial execution is resource intensive, presenting numerous logistical challenges. First, human resources are needed, often requiring large sample sizes depending on the number of arms, hypothesis being tested and recruitment time. For example, if there is a standard treatment that has an efficacy rate of 0.60, and the new treatment is expected to have an efficacy rate of 0.70. The sample size required under the general conditions of 5% of α -error and 20% of β -error to show that the new treatment is statistically significantly more effective is 356 patients in each group. Secondly, time-dependent endpoints, such as progression-free survival (PFS) and overall survival (OS) may require years to reach the critical number of events for interpretation. Increasingly, co-primary endpoints of PFS and OS are being evaluated, which not only extends this time to report but can over-power the proximal endpoint if, as is often the case for ovarian cancer trials, the post-progression survival expectation is long.

Third, large trials require substantial funding. The Johns Hopkins Bloomberg School of Public Health published that 'Clinical trials that support FDA approvals of new drugs have a median cost of \$19 million [3]. The \$19 million median figure represents less than one percent of the average total cost of developing a new drug, which in recent years has been estimated at between \$2 to \$3 billion [4].

Therefore, an interim analysis was devised to determine whether a decision-making conclusion on efficacy or safety could be reached before the end of a clinical trial, according to the method specified in the protocol. A clinical trial is designed by calculating the sample size required for statistical analysis based on the previously obtained results of each arm's treatment and the defined values of α - and β -errors. To determine the statistical significance of the primary endpoint, the Lan-DeMets α -spending function is typically used to adjust for the multiplicity of tests [5]. There are two approaches to the α -spending function: the Pocock approach [6] and the O'Brien-Fleming approach [7]. The reason for using the α -spending function is that although the log-rank test results depend on the number of patients enrolled, the number of enrolled patients does not reach the target sample size at the time of an interim analysis. To determine whether to continue or discontinue a clinical trial based on a statistically significant difference in the primary endpoint, a method is used to compare the significance of the p-values at a certain time point with a sample size smaller than the sample size with the α -error value at a predetermined final time point, usually 0.05. In other words, the decision to continue or discontinue a clinical trial depends on the statistic. Interim analysis is also used to determine the futility of discontinuing a clinical trial early because it is predicted that it will be difficult to

demonstrate the efficacy of the trial treatment if the trial continues. If the significance of the p-values at a certain time point with a sample size is larger than the α -spending function and the conditional power is larger than a certain value, the trial should be discontinued as a futility stop [8-10].

If the lacking patient data can be predicted and supplemented by computation with good reliability, the α -spending function might become unnecessary, and human, time and financial resources can be saved. Therefore, to complement the lacking data, we have developed and published a novel strategy to predict unknown data from a smaller sample size by using compressive sensing [11-14] which is one of the information technologies used in artificial intelligence [15,16]. This method generates a complete list of the predicted data from a smaller number of known data. Unlike the current interim analysis, this method (a) uses AI that can classify data distributions of not only the known uncensored data but also the known censored data, respectively, (b) can generate predicting time data by compressive sensing, (c) can generate thousands of p-values calculated by a surviving test such as a log-rank test resulting in robust conclusion presumably, and (d) would be able to predict the outcome of survival in the form of p-value distribution profile any time during the trial. By this method, the number of patients reaches the target sample size, so the precondition of α - and β -errors is retained, and α -errors can be simply compared without introducing the concept of α -spending function.

In this study, we demonstrated this new method for interim analysis to obtain a p-value distribution profile of the log-rank test at various times before a clinical trial would be finished. This is by iteratively predicting the lacking time data of censored and uncensored cases respectively for each arm at the time when the determined sample size will be achieved using an improved algorithm for comparing two groups of clinical trial data from a real world database in Vivli Inc. (<https://vivli.org>). Data provided was deidentified individual participant-level data (IPD) [17] that used the raw unit-level data from each primary study [17,18] and allowed different sources of heterogeneity in the effect estimate to be explored [19]. We investigated how much our method predicts the final p-value of log-rank test and how accurate the decision for whether trial continuation with fewer sample size. The feasibility of this new prediction method as a new interim analysis was investigated using the IPD.

MATERIALS AND METHODS

Targeted Database

We used databases of clinical trials in Vivli Inc. (<https://vivli.org>). For this study, we selected five clinical trials that were two-arm comparisons, enrolled at least 300 patients, and showed Kaplan-Meier curves and log-rank test results for PFS and OS, respectively, in the published papers. The selected five clinical trials were as follows; Study No.1: A Double-blind, Randomised, Placebo Controlled Phase III Study of Nintedanib Plus Best Supportive Care (BSC) Versus Placebo Plus BSC in Patients with Metastatic Colorectal Cancer Refractory to Standard Therapies. [20], Study No.2: A Randomized, Double-blind, Multicenter Phase 3 Study of Irinotecan, Folinic Acid, and 5-Fluorouracil (FOLFIRI) Plus Ramucirumab or Placebo in Patients With Metastatic Colorectal Carcinoma Progressive During or Following First-Line Combination Therapy With Bevacizumab, Oxaliplatin, and a Fluoropyrimidine [21], Study No.3: A Phase III, Multicenter, Randomized, Blinded, Placebo-controlled Trial of Carboplatin and Gemcitabine Plus Bevacizumab in Patients With Platinum-sensitive Recurrent Ovary, Primary Peritoneal, or Fallopian Tube Carcinoma [22], Study No.4: AURELIA: A Multi-center, Open-label, Randomised,

Two-arm Phase III Trial of the Effect on Progression Free Survival of Bevacizumab Plus Chemotherapy Versus Chemotherapy Alone in Patients With Platinum-resistant, Epithelial Ovarian, Fallopian Tube or Primary Peritoneal Cancer for PFS [23] and OS [24], Study No.5: A Randomized, Open-label Phase III Intergroup Study: Effect of Adding Bevacizumab to Cross Over Fluoropyrimidine Based Chemotherapy (CTx) in Patients With Metastatic Colorectal Cancer and Disease Progression Under First-line Standard CTx/Bevacizumab Combination [25].

We remotely logged into the research environment of Vivli Inc. and obtained deidentified individual participant-level data of the studies for each arm for PFS and OS, including the date of randomization, the last event for censored or uncensored, and the date of the last event confirmation.

The data was processed at the date of analysis so that the longest observation period might not exceed the observation period. For example, if the observation period is set to 5 years, a case that was not censored at 5.1 years was considered a censored case at 5 years, and a case that was censored at 5.1 years was considered a censored case at 5 years. All analyses will use this process for all patient's data with each appropriate observation date as determined. In this analysis, the observation period was defined as 5 years that corresponds to 1,826.25 ($=365.25 \times 5$) days. Then each study was re-evaluated by log-rank test with processed data. Due to this process, the results of the analysis by this study might differ from the results already published in the corresponding papers.

Because the date information was not found in the database for Study No.5, it was excluded from the analysis, and the results were obtained from the remaining four studies. The PFS/OS in study 1, 2, 3 and 4, was defined as trial No.1/No.2, No.3/No.4, No.5/No.6 and No.7/No.8, respectively.

Prediction of the Number of Uncensored Cases in the Final Sample Size

In the log-rank test, the number of censored cases is irrelevant to the prognostic calculation and the p-value, and the number of uncensored cases affects the p-value. Hence, it is important to predict the number of uncensored cases at the final sample size. The ratio of the number of samples at a certain point in time to the final sample size is called the information fraction, and the date corresponding to the appropriate information fraction at which the number of events was ten or more was determined because the surviving analysis with log-rank test is not reliable with a few samples. The classification of censored or uncensored was confirmed by the date for each case and varied depending on the date. As a common interim analysis for comparison, the p-values of the Lan-DeMets spending function as determined by the information fraction to approximate an O'Brien-Fleming bound in which $\alpha=0.025$ for one-sided Type I error specification and the conditional powers were calculated.

The list in the ratio of uncensored cases at appropriate date intervals, which was one month in this study, up to the date just determined was obtained. From the product of the known final sample size and the proportion of uncensored cases corresponding to the 5th, 25th, 50th, 75th, and 95th percentiles in this list, we obtained five different estimates of the number of uncensored and censored cases, respectively.

Compressive Sensing

Next, we set the appropriate dates where the number of events, or the number of uncensored events, is ten or more. Then, data of the corresponding information fractions and a matrix of random elements composed of the optimal distribution function for the data for censored and uncensored data were obtained, respectively. The functional forms of distributions as candidates for data distributions that were the same as we had previously reported for analyzing cases of massive hemorrhage during delivery [26] were as follows; beta distribution [27,28], Cauchy distribution [27,29], chi distribution [30], chi-square distribution [27,31], exponential distribution [27,32], extreme value distribution [27,33], Frechet distribution [34], gamma distribution [35], Gumbel distribution [36], half normal distribution [37], inverse Gaussian distribution [27,38,39], Laplace distribution [27,40], Levy distribution [41], logistic distribution [27,42,43], log normal distribution [27,38,44,45], Maxwell distribution, normal distribution [27,44], Pareto distribution [44,46], Rayleigh distribution [27,47], Student-t distribution [27,48], uniform distribution [49], Weibull distribution [27] and mixed distribution consisted of more than two distributions above.

Our method of compressive sensing was used to produce five different data sets for data completion. The data of the remaining patients who should be enrolled will be predicted as follows:

$$\arg \min \frac{1}{2n} \|Y - AX\|^2 + \lambda \sum_{j=1}^m |x_j|$$

where m ; the number of sample size of patients to fit a clinical trial, n ; enrolled patients, $\lambda \in \mathbb{R}$, $X = {}^t(x_1, \dots, x_m)$, $Y = {}^t(y_1, \dots, y_n)$, $m, n \in \mathbb{N}$, $m > n$, $\forall x \geq 0$, $\forall y \geq 0$, $\forall S \in$ data distribution of Y , $A \in S_{n \times m}$ and $\forall y \in X$, as we reported [15,16].

To increase the reliability of prediction by compressive sensing as a computational technique, we added nine times as many zeros as the number of cases to the censored and uncensored data, respectively, to sparse the data, because compressive sensing can reconstruct data well especially when data are sparse [15,16]. Eventually, the added zeros were excluded to obtain a data set with the target sample size. The data set can be created any number of repeated times, but in this case, three times was chosen. In this way, we obtained 15 (= 5×3) prediction data for each group.

Log-Rank Test

We then performed 225 (= (5×3)²) log-rank tests for all combinations of the two groups; arm A and B, at some appropriate information fractions that ranges universally from 0.2-0.3 to about 1.0. Assuming that the proportion of uncensored cases during the observation period follows a standard normal distribution $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, the data corresponding to the 5th, 25th, 50th, 75th, and 95th percentiles were obtained with rounded down to two decimal places. We assumed that the data corresponding to a percentile value have a weight that was $f(\tau)$, where $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $v = \int_{-\infty}^{\tau} f(x)$ and v was the percentile value. Thus, the data corresponding to the 5th, 25th, 50th, 75th, and 95th percentiles have weights of 0.10, 0.31, 0.39, 0.37, and 0.10 of the

cumulative distribution function value of the standard normal distribution, respectively, so we can multiply the number of log-rank test p-values by 10, 31, 39, 31, and 10, respectively, so that 819,675 p-values of log-rank test for each information fraction might be obtained. Statistical values such as median, mean, 5th percentile, 95th percentile, and standard deviation of each set of log-rank p-values were obtained.

Evaluation of Log-Rank P-Value Distribution

The following evaluation criteria were defined to determine whether to continue the clinical trial of this method.

The first method to the evaluation was that if the 95th percentile of the predicted log-rank p-values was less than 0.05, it was expected that the clinical trial would show significant differences and the current clinical trial should be discontinued as efficacy stop. If the 5th percentile of the predicted log-rank p-values was equal to or more than 0.05 and the power was equal to or more than 0.2 [8], it is expected that the clinical trial will show no significant differences and the current clinical trial should be discontinued as futility stop. If the 5th percentile of the predicted log-rank p-values was equal to or more than 0.05 and the power was less than 0.2, the current clinical trial should be continued, though the trial could be considered worthy of discontinuation as a futility stop without statistical confidence. In other cases, such as when the median value was less than 0.05 but the 95th percentile value was greater than 0.05, the decision was continuation.

As for the conventional interim analysis, the criterion of the conditional power was defined as 0.2 [8] in this study. The power and the conditional power were compared by t-test when information fraction was less than 1.00.

Development Environment of Program and Statistical Analysis

An original program for this study was created with using Mathematica 12.3 (Wolfram Research, Champaign, IL, USA) at Medical Data Labo (Okayama, Japan). The program was transferred to Windows server (Microsoft, Redmond, WA, USA) with Xenon 2.6 GHz and 14 GB RAM in Research Environment (Vivli Inc.) where all of processes for analysis were carried out remotely with Microsoft Remote Desktop 10.7.1 (Microsoft, Redmond, WA, USA) via the Internet. Log-rank test and t-test were used with using Mathematica 12.3. $P < 0.05$ was considered to indicate statistical significance.

RESULTS

The p-values of log-rank test for PFS, OS and the number of finally enrolled patients that were recalculated for study 1, 2, 3 and 4 were 9.351×10^{-12} , 0.41185 and 814, 0.00194, 0.04882 and 768, 1.8357×10^{-11} , 0.34293 and 484, and 3.2349×10^{-12} , 0.22017 and 361, respectively (Table 1).

As for trial No.1 with 9.351×10^{-12} of log-rank p-value, the ratio of uncensored cases of arm A as a function of information fraction gradually increased until the information fraction reached 0.4, and then it remained at almost the final ratio 0.4. The ratio of arm B gradually increased until the information fraction reached 0.4, and then it remained at almost the final ratio 0.6. From the early stage of the information fraction of 0.3256, the final predicted value by our method was expected to be below 0.05 and correct (Figure 1). Our method was able to make

the right decision as efficacy stop earlier than the conventional method (Table 2). As for trial No.2 with 0.41185 of log-rank p-value, the ratio of uncensored cases of arm A and arm B as a function of information fraction ratio was constantly increasing until the information fraction reached 1.0. The final values were about 0.2. The log-rank test could not be performed because few events occurred when the information fraction was less than 0.4245. From the early stage of the information, the final predicted value by our method was expected to be over 0.05 and was correct (Figure 2). In our method, the judgement when information fractions were equal to or more than 0.6068 might be worthy of futility stop because 5th percentile were more than 0.05. However, since powers were less than 0.2, the judgement was a continuation. The conventional method resulted in a continuation because the conditional powers were <0.2 . The judgements of our method and the conventional method were similar. (Table 3). As for trial No.3 with 0.00194 of log-rank p-value, the ratio of uncensored cases of both arm A and B as a function of information fraction ratio was constantly increasing to 0.8 until the information fraction reached 1.0. From the early stage of the information fraction of 0.1733, the final predicted value by our method was expected to be below 0.05 and was correct (Figure 3). Our method was able to make the right decision as efficacy stop earlier than the conventional method. (Table 4). As for trial No.4 with 0.04882 of log-rank p-value, the ratios of uncensored cases of both arm A and arm B as a function of information fraction were constantly increasing to about 0.4 until the information fraction reached 1.0, then increased to 1.0. The 95th percentiles, median and mean of the final predicted values were greater than or less than 0.5 and the 5th percentiles of the values were less than 0.05. The 95th percentile of the final predicted value by our method was less than 0.05 only at information fraction at 0.4320 (Figure 4). The judgements of our method and the conventional method were similar (Table 5). As for trial No.5 with 0.8357×10^{-11} of log-rank p-value, the ratio of uncensored cases of arm A as a function of information fraction ratio was constantly increasing to 0.5 until the information fraction reached 1.0. The final value was about 0.75. The ratio of arm B was constantly increasing to 0.6 until the information fraction reached 1.0. The final value was about 0.9. From the information fraction of 0.4811, the final predicted value by our method was expected to be below 0.05 and correct (Figure 5). The judgements of our method and the conventional method were similar (Table 6). As for trial No.6 with 0.34293 of log-rank p-value, the ratio of uncensored cases of arm A as a function of information fraction ratio was constantly increasing to about 0.1, which was a low value, until the information fraction reached 1.0. The final value increased to about 0.7. The ratio of arm B was constantly increasing to about 0.2, which was also a low value, until the information fraction reached 1.0. The final value increased to about 0.7. After the planned sample size was achieved, the value of the log-rank test increased significantly with the passage of observation time while the information fraction remained at 1.0. The log-rank test could not be performed until the information fraction was less than 0.7735 because few events occurred. From the information fraction of 0.8226 to close to 1.0, the final predicted value by our method was expected to be below 0.05 and was incorrect (Figure 6). The judgements of our method and the conventional method were similar. (Table 7). As for trial No.7 with 3.2349×10^{-12} of log-rank p-value, the ratio of uncensored cases of arm A as a function of information fraction increased to 0.5 until the information fraction reached 0.4, and then it constantly and gradually increasing to about 0.7 until the information fraction reached 1.0. The final value increased to 1.0. The ratio of arm B was constantly increasing to about 0.6 until the information fraction reached 1.0. The final value increased to 1.0. From the information fraction of 0.2528, the final predicted value by our method was expected to be below 0.05 and correct (Figure 7). Our method was able to make the right decision as efficacy

stop earlier than the conventional method. (Table 8). As for trial No.8 with 0.22017 of log-rank p-value, the ratio of uncensored cases of arm A as a function of information fraction gradually increased to about 0.2 until the information fraction reached 1.0. The final value increased to about 0.7. The ratio of arm B gradually increased to about 0.2 until the information fraction reached 1.0. The final value increased to about 0.8. The 95th percentiles, median and mean of the final predicted values value by our method was greater than or less than 0.05. The 95th percentiles of the final predicted values were less than 0.05 between 0.5306 and 0.7167 of information fraction (Figure 8). Our method was not able to make the right decision that was futility stop earlier than the conventional method. The judgements of our method and the conventional method were similar. (Table 9).

The number of cases judged to have been correctly efficacy stop, correctly futility stop, wrong efficacy stop, wrong futility stop and remaining continuation to the end by our method for significant/non-significant trial results were 4/NA, NA/0, 0/1, 0/0 and 1/2, respectively. The number of cases for those by the conventional method were the same. The values of power were larger than the values of the conditional power in all trials when information fraction was less than 1.00.

The effect of the number of calculations on the results was examined, and the results for both the log-rank p-values and the value of the power remained almost the same even after more than one million calculations for each information fractions (data not shown).

DISCUSSIONS

The final p-values of log-rank test could be predicted to some extent by iteratively predicting the lacking time data using our method for comparing two groups of clinical trial data. We thought one major reason was that the data used in this study were not such that it would have resulted in a futility stop in the first place, and there was no difference between our method and the conventional method in that data set. But we considered it statistically important that we had developed a method that was at least comparable to the conventional method. It is impossible to predict the future completely, but we developed a new method derived from compressive sensing that is one of the information technologies used in artificial intelligence that we have applied to medical studies in recent several years [15,16,26,50-61]. Our method for predicting unknown data including the observed data can be applied to various types of data including length, weight, volume etc. In this study, time associated with date in clinical trial was targeted. We applied the method for interim analysis to estimate the final prediction of survival from a small number of enrolled patients, which is a completely different approach from the method of the Lan-DeMets α -spending function [5]. Since our method complements data of the lacking cases, we do not need to introduce the concept of the spending function because our method satisfies the error condition that was already required in the clinical trial design. The results predicted by our method does not present a confirmed outcome but data distributions of the probability of log-rank test. Because the single outcome is only a candidate of unknown data, robustness of the result could be obtained by repeated calculations. In an empirical case of our previous report, we showed that the p-value distribution of the log-rank test for thousands or millions can be obtained by simply iterating the missing data [15,16]. However, to speed up the calculation and to incorporate errors in the estimated number of uncensored and censored cases, we used percentile values and corresponding weights from standard normal distribution in this study. The confidence intervals could be obtained and the

calculating time became more than one hundred times faster with the process (data not shown). Prediction by this method can be repeatedly obtained at various information fractions in the clinical trial when the information of a single patient is censored.

As for technical view of computer calculation, compressive sensing can reconstruct data better especially when data are sparse. Thus, we added nine times zeros to the known data, obtained the shortage data, and removed the zeros. we thought it was not necessary for investigators to analyze and investigate the sparseness of enrolled data by the procedure of adding and removing zeros. Additionally, compressive sensing requires a matrix $A \in S_{n \times m}$ consisted of some data distribution function. To make a matrix for each enrolled data at an information fraction, we identified and used data distributions such as Weibull distribution, log normal distribution and so on. We thought this method to make a matrix was very important to complete lacking data. Though this calculation process might need further investigation, we thought our method demonstrated in this study was acceptable because of good performance.

With using quantitative criteria, all final p-value predictions for PFS by our method were correct, probably because event occurrence was observed relatively earlier compared to OS.

When α -error by log-rank test was focused, our method seemed to have made the right decision earlier than the Lan-DeMets method in deciding whether to continue or discontinue the trial. When the final actual log-rank p-value was quite small, as was the case for all four PFS trials, even when the information fraction was around 0.2-0.4, the final p-value predicted at that time by our method was less than 0.01. Therefore, if the predicted final p-value by our method at the information fraction of 0.2-0.4 was about 0.05, it could be decided that the trial should be discontinued as efficacy stop if the power was more than 0.2. The predicted final p-value was correct even at an early stage, such as an information fraction of about 0.2, when more than ten events occurred. On the other hand, the predicted final p-value was not correct when the number of events occurred in less than ten patients even though the number of enrolled patients was large, as in the case of Trial No. 6, where the information fraction was 0.9, and when events occurred later in the observation period. In the case of clinical trials in which events did not occur frequently, such as Trial No. 2 and 6, the number of events was too small to calculate well even if the information fraction was quite high. In cases such as Trial No. 4, 6, 7 and 8, where the event occurred gradually after the enrollment was completed and observation was continued, the p-value increased from less than 0.05 to 0.2, etc. while the information fraction remained at 1.0. The final p-value could be difficult to predict in these cases. If the final predicted p-value was around 0.05 such as trial No. 4 of which p-value of the actual result was 0.0482, it was difficult to predict the final significance.

The Kaplan-Meier curve should not be crossed in survival analysis and is of clinical and statistical use. The curve crossing was observed in trial No. 2 but did not affect the final p-value prediction. The time course of survival could be assumed to be the occurrence of an event according to a specific probability distribution, and our method does not require the assumption of such an unknown function and uses the distribution of the raw data as it is. Therefore, it is considered that the results of log-rank test might not be affected even if the curve was crossed.

There are some limitations of our method. First, In the case of a clinical trial in which events did not occur easily, the number of events was too small to calculate the survival rate well even if the information fraction was quite advanced. These might be due to the following two reasons. Since the log-rank test in survival analysis is greatly affected by the number of events, especially at the time when the observation time is long, if the number of events itself is small, a small change in the number of events can cause a large change in the p-value. Using little data, the prediction of the data distribution function in our method will be more inaccurate, so the error of the predicted data will be larger. The temporal status of event occurrence is generally unknown before the study begins. Therefore, if fewer than ten events occurred in at least one arm, it was considered prudent to apply this method and to continue the clinical trial. As seen in Trial No.2, if the predicted final p-value at an information fraction of about 0.4 in the case of a clinical trial with a low event frequency throughout the entire observation period was already high enough such as about more than 0.5, we thought that the trial should be discontinued as futility stop if the power was more than 0.2. Second, if the information fraction remains at 1.0 and the final p-value increased from less than 0.05 to 0.2 due to the gradual occurrence of events with no new enrollment, and observation was continued after the expected number of sample size was achieved, the final prediction was difficult, as seen in Trial No. 6 (Figure 6). We considered this situation to occur if the event was a gradual occurrence or if the observation period was long relative to the time leading up to the event. In clinical trials, if the final predicted p-value is lower than 0.05 when the information fraction is small and the power is less than 0.2, the clinical trial should be continued. Incidentally, we can predict that if we further increase the number of required sample size, the prediction will become more accurate, but this is not a good solution as it goes against the purpose of improving the efficiency of clinical trials. Third, it would be difficult to predict the judgement whether significant or not in cases where the final actual predicted p-value was around 0.05, such as trial No. 4 of which the final p-value was 0.0488. Fourth, the appropriate observation time that was arbitrarily defined as 1826.25 days in this study was unknown. Because we did not want to expect infinite censoring, this finite time was defined as the observation time. The appropriate observation time seemed to be dependent on the incidence of how often the event would occur. If the incidence of event was less than 0.2, it might be better to define the observation time more than 1826.25 days. This appropriateness of selected observation time might be considered in the study design protocol for interim analysis using information from preliminary trials for not only our method but also the conventional method.

In clinical trials, we generally expect to see significant differences, and it is preferable to not make easy aborts at an early stage with cautious judgment. Therefore, the calculated power was used to avoid futility stop with confidence as much as possible. There were no trials determined to be futility stop in the studies, because all the studies were not discontinued but completed, and the values of power by our method were larger than the values of conditional power using the conventional method. Therefore, if the abandoned clinical trial data was investigated by our method, our method might perform better due to the larger values of power. Our method could predict the evaluation of the judgement for trials with statistically significant results correctly. In all the clinical trials with PFS as the endpoint, our method correctly predicted that there were final significant differences and sufficient powers from the early stage of the information fraction, so efficacy stop was recommended.

Although the present study was analyzed assuming a two-stage design, our method for complementing lacking data is expected to be available for the adaptive group sequential design that allows the researcher to represent every group sequential plan as an adaptive trial design and to make design changes during the course of the trial after every interim analysis in the same way as with adaptive designs [62-66], and adaptive seamless design that combines a phase II and phase III into one trial incorporating a decision to adapt the study at an interim analysis and allows modifications made to trial and/or statistical procedures of ongoing clinical trials [67-70]. Our method that is not constrained by the α -spending function might be valuable for validating the feasibility and applicability of not only two-stage study design but also those adaptive designs.

Our method using compressive sensing, an information technology of artificial intelligence, is expected to be useful as a new interim analysis with date, focusing on changes in the number and the rate of event occurrences. Our method was at least comparable to the conventional method. It may be possible to expand the options of applicable methods for interim analysis. Our method may be worth applying to future trials that are not terminated futilely but are not significant at the end. We can expect that there will be trials that could have been stopped earlier if our method had been used. By supplementing the data on the number of lacking patients by our method, it will be possible to improve the efficiency of human, time, and financial resources. It would be helpful in establishing efficacy and futility judgement in clinical trials.

Contributor Roles Taxonomy Author Statement

Yasunari Miyagi: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision. **Keiichi Fujiwara:** Validation, Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration. **Hisanaga Nomura:** Data Curation, Validation, Writing - Review & Editing. **Koji Yamamoto:** Validation, Investigation, Writing - Review & Editing, Supervision. **Robert L Coleman:** Validation, Writing - Review & Editing, Resources, Review & Editing, Supervision.

All the authors confirm they had full access to all the data in the study and they accept the responsibility to submit for publication. All the authors approved the final version of the article.

Conflict of Interest Statements

The authors declare no conflict of interest.

TABLES

Table 1. The summary of the trials. Each trial was re-evaluated by log-rank test for a defined observation period. The data was processed at the date of analysis so that the longest observation period does not exceed 1826.25 days (= 5 years). Due to this process, the recalculated p-value of log-rank test in this study may differ from the p-values already published in the corresponding paper. OS; overall survival, PFS; progression-free survival.

Study No.	Clinical trial No.	Cancer	Arm design	The number of enrolled patients	Endpoint	Recalculated P-Value of Log-Rank Test	Original P-Value of Log-Rank Test
Study No.1 [20]	Trial 1	Metastatic colorectal carcinoma	nintedanib versus placebo	814	PFS	9.351×10^{-12}	<0.0001
	Trial 2				OS	0.411	0.8659
Study No.2 [21]	Trial 3	Metastatic colorectal carcinoma	ramucirumab plus FOLFIRI versus placebo plus FOLFIRI	768	PFS	0.001936	0.0005
	Trial 4				OS	0.048818	0.0219
Study No. 3 [22]	Trial 5	Platinum-sensitive Recurrent Ovary, Primary Peritoneal, or Fallopian Tube Carcinoma	carboplatin and gemcitabine plus bevacizumab versus carboplatin and gemcitabine plus placebo	484	PFS	1.836×10^{-11}	<0.0001
	Trial 6				OS	0.3429	0.6479
Study No. 4	Trial 7 [23]	Platinum-resistant, Epithelial Ovarian, Fallopian Tube or Primary Peritoneal Cancer	bevacizumab plus chemotherapy versus chemotherapy alone	361	PFS	3.235×10^{-12}	<0.001
	Trial 8 [24]				OS	0.2202	<0.174

Table 2. Result for PFS of trial No.1 [20] with the actual final p-value of log-rank test was 9.351×10^{-12} . From the early stage of the information fraction of 0.3256, the 95th percentiles of the final predicted values by our method were less than 0.05. The judgement was efficacy stop at this stage. On the other hand, this trial would have been suggested to continue at the early stage until the information fraction reached 0.4905 by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. In the conventional method, the judgement when information fraction was equal or less than 0.3256 was either futility stop with no statistical confidence that conditional power was lower than 0.2 or continuation. Our method was able to make the right decision as efficacy stop earlier than the conventional method. The values of power were larger than the values of the conditional power ($P=0.039$ by t-test).

Information fraction	P-value of log-rank test at the information fraction	Predicted p-value of log-rank test at end of the clinical trial					Power	Judgement by our method	Conventional method		
		Median	Mean	Standard deviation	5 th percentile	95 th percentile			P-value [†]	Conditional power	Judgement [‡]
0.1703	4.968×10 ⁻³	2.650×10 ⁻⁵³	1.616×10 ⁻⁵⁰	3.560×10 ⁻⁵⁰	1.408×10 ⁻⁵⁴	1.050×10 ⁻⁴⁹	0.999	efficacy stop	5.576×10 ⁻⁸	0.112	continuation [§]
0.2662	1.612×10 ⁻³	4.898×10 ⁻⁷	0.00825	0.2312	1.982×10 ⁻²³	0.6690	0.702	continuation	1.398×10 ⁻⁵	0.160	continuation [§]
0.3256	2.913×10 ⁻³	3.997×10 ⁻⁸	0.01290	0.07039	1.006×10 ⁻¹⁵	0.01380	0.832	efficacy stop	8.580×10 ⁻⁶	0.168	continuation [§]
0.4905	4.765×10 ⁻⁶	6.910×10 ⁻⁸	1.259×10 ⁻⁴	4.054×10 ⁻⁴	1.348×10 ⁻⁹	6.790×10 ⁻⁴	0.831	efficacy stop	1.373×10 ⁻³	0.394	efficacy stop
0.6608	2.825×10 ⁻⁹	3.753×10 ⁻¹³	1.550×10 ⁻¹²	3.176×10 ⁻¹²	1.200×10 ⁻¹⁴	7.644×10 ⁻¹²	0.978	efficacy stop	5.828×10 ⁻³	0.710	efficacy stop
0.8027	1.052×10 ⁻⁹	1.010×10 ⁻¹²	1.164×10 ⁻¹¹	5.010×10 ⁻¹¹	1.110×10 ⁻¹³	5.184×10 ⁻¹¹	0.896	efficacy stop	0.01235	0.762	efficacy stop
0.9770	3.262×10 ⁻¹²	3.570E-12	6.330×10 ⁻¹²	8.285×10 ⁻¹²	1.310×10 ⁻¹²	2.202×10 ⁻¹¹	0.867	efficacy stop	0.02335	0.884	efficacy stop
0.9980	1.714×10 ⁻¹¹	1.458×10 ⁻¹¹	1.435×10 ⁻¹¹	1.613×10 ⁻¹²	1.238×10 ⁻¹¹	1.714×10 ⁻¹¹	0.661	efficacy stop	0.02490	0.908	efficacy stop

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 3. Result for OS of trial No.1 [20] with the actual final p-value of log-rank test was 0.41185. The judgement was continuation from the early stage. This trial would have been suggested to continue at the early stage by using either our method or the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. In our method, the judgement when information fractions were equal to or more than 0.6068 might be worthy of futility stop because 5th percentile were more than 0.05. However, since powers were less than 0.2, the judgement was a continuation. The conventional method resulted in a continuation because the conditional powers were <0.2. The judgement of our method and the Lan-DeMets spending function were similar. The judgement of our method and the conventional method were similar. The values of power were larger than the values of the conditional power (P=0.006 by t-test).

Information fraction	P-value of log-rank test at the information fraction	Predicted p-value of log-rank test at end of the clinical trial					Power	Judgement by our method	Conventional method		
		Median	Mean	Standard deviation	5 th percentile	95 th percentile			P-value [†]	Conditional power	Judgement [‡]
0.4245	0.5143	0.5023	0.4802	0.1513	0.0394	0.7061	0.106	continuation	0.0006	0.012	continuation [§]
0.5156	0.8645	0.3460	0.4104	0.3268	0.0220	0.9510	0.238	continuation	0.0018	0.004	continuation [§]
0.6068	0.4853	0.6426	0.6437	0.1591	0.3300	0.9000	0.066	continuation [§]	0.0040	0.013	continuation [§]
0.6979	0.8966	0.6456	0.6468	0.1592	0.3576	0.8685	0.075	continuation [§]	0.0073	0.004	continuation [§]
0.8398	0.2207	0.1898	0.2308	0.1139	0.0683	0.5766	0.175	continuation [§]	0.0145	0.029	continuation [§]
0.9987	0.2692	0.2636	0.2615	0.0063	0.2520	0.2690	0.148	continuation [§]	0.0249	0.024	continuation [§]

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 4. Result for PFS of trial No.2 [21] with the actual final p-value of log-rank test was 0.00194. From the early stage of the information fraction of 0.1733 except 0.6738, the 95th percentiles of the final predicted values are less than 0.05 so that the trial might be correctly efficacy stop. When the information fraction of our method was 0.1155 and 0.6738, 95th percentile was more than 0.05 though the median value and the mean value were less than 0.05 and the judgement resulted in continuation. On the other hand, this trial would have been suggested to continue until the information fraction reached 0.8642 by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. In the conventional method, all cases with information fraction ≤ 0.6738 were futility stops, which were all misjudgments, but the small conditional power resulted in continuation. Our method was able to make the right decision as efficacy stop earlier than the conventional method. The values of power were larger than the values of the conditional power ($P=0.0039$ by t-test).

		Predicted p-value of log-rank test at end of the clinical trial							Conventional method		
Information fraction	P-value of log-rank test at the information fraction	Median	Mean	Standard deviation	5 th percentile	95 th percentile	Power	Judgement by our method	P-value [†]	Conditional power	Judgement [‡]
0.1155	0.08892	4.230×10 ⁻⁸	0.03060	0.1263	2.080×10 ⁻¹⁹	0.1357	0.804	continuation	4.252×10 ⁻¹¹	0.053	continuation [§]
0.1733	0.03211	6.528×10 ⁻⁷	3.930×10 ⁻³	0.01568	2.715×10 ⁻¹⁷	0.01637	0.644	efficacy stop	7.250×10 ⁻⁸	0.083	continuation [§]
0.2374	8.801×10 ⁻³	1.738×10 ⁻⁸	4.523×10 ⁻⁵	3.405×10 ⁻⁴	4.466×10 ⁻¹²	9.563×10 ⁻⁵	0.878	efficacy stop	4.226×10 ⁻⁶	0.134	continuation [§]
0.3208	9.919×10 ⁻³	1.350×10 ⁻⁵	4.635×10 ⁻⁴	2.520×10 ⁻³	1.782×10 ⁻⁹	1.769×10 ⁻³	0.640	efficacy stop	7.590×10 ⁻⁵	0.133	continuation [§]
0.507	0.01901	8.300×10 ⁻⁴	6.440×10 ⁻³	0.01490	6.301×10 ⁻⁶	0.03026	0.464	efficacy stop	1.640×10 ⁻³	0.115	continuation [§]
0.6738	0.01242	4.410×10 ⁻³	0.05319	0.1344	1.997×10 ⁻⁶	0.2623	0.328	continuation	6.322×10 ⁻³	0.142	continuation [§]
0.8642	1.817×10 ⁻³	1.468×10 ⁻³	4.807×10 ⁻³	8.520×10 ⁻³	8.712×10 ⁻⁵	0.02233	0.226	efficacy stop	0.01590	0.249	efficacy stop
0.9197	5.753×10 ⁻⁴	5.680×10 ⁻⁴	1.150×10 ⁻³	2.123×10 ⁻³	1.996×10 ⁻⁵	2.990×10 ⁻³	0.221	efficacy stop	0.01940	0.321	efficacy stop

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 5. Result for OS of trial No.2 [21] with the actual final p-value of log-rank test was 0.04882. The 95th percentiles of the final predicted values by our method straddled the top and bottom across 0.05. The judgement was continuation at the most of information fractions. This trial would have been suggested to continue at the any stage by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. The conventional method resulted in a continuation as a judgement because the conditional power was <0.2, though the judgement was a futility stop by considering P-value only. The judgement of our method and of the conventional method were similar. The values of power were larger than the values of the conditional power (P=0.004 by t-test).

Information fraction	P-value of log-rank test at the information fraction	Predicted p-value of log-rank test at end of the clinical trial					Power	Judgement by our method	Conventional method		
		Median	Mean	Standard deviation	5 th percentile	95 th percentile			P-value [†]	Conditional power	Judgement [‡]
0.1166	0.2887	0.1150	0.0939	0.0809	1.535×10 ⁻⁸	0.2279	0.517	continuation	5.215×10 ⁻¹¹	0.028	continuation [§]
0.2406	0.2755	0.0040	0.0417	0.0760	5.558×10 ⁻⁶	0.2250	0.623	continuation	4.897×10 ⁻⁶	0.025	continuation [§]
0.4320	0.0292	0.0014	0.0030	0.0044	1.731×10 ⁻⁵	0.0124	0.780	efficacy stop	6.490×10 ⁻⁴	0.078	continuation [§]
0.5957	0.1072	0.0115	0.0906	0.1791	5.175×10 ⁻⁵	0.4488	0.469	continuation	3.684×10 ⁻³	0.048	continuation [§]
0.7733	0.0361	0.0718	0.1401	0.1910	9.633×10 ⁻⁴	0.5446	0.359	continuation	0.01806	0.086	continuation [§]
0.9390	0.0635	0.0434	0.0457	0.0123	0.03012	0.0779	0.156	continuation	0.02072	0.068	continuation [§]

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 6. Result for PFS of trial No.3 [22] with the actual final p-value of log-rank test was 1.8357×10^{-11} . From the information fraction of 0.4811, the 95th percentiles of the final predicted values are less than 0.05. The judgement was efficacy stop at the information fraction of 0.4811. Similarly, this trial would have been suggested to efficacy stop until the information fraction reached 0.4811 by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. In the conventional method, all cases with information fraction ≤ 0.4328 were futility stops, but the small conditional power resulted in continuation. The judgement of our method and of the conventional method were similar. The values of power were larger than the values of the conditional power when information fraction was less than 1.00 ($P=0.011$ by t-test).

Informa tion fraction	P-value of log- rank test at the informa tion fraction	Predicted p-value of log-rank test at end of the clinical trial					Pow er	Judgeme nt by our method	Conventional method		
		Media n	Mean	Standar d deviati on	5 th percen tile	95 th percen tile			P- value [†]	Condi tional power	Judgement [‡]
0.2983	0.3791	1.493×10^{-4}	1.394×10^{-3}	3.070×10^{-3}	1.696×10^{-6}	5.923×10^{-3}	0.343	efficacy stop	4.065×10^{-5}	0.020	continuation [§]
0.3887	0.06152	0.08711	0.2146	0.2379	9.780×10^{-4}	0.7292	0.352	continuation	3.240×10^{-4}	0.06	continuation [§]
0.4055	0.03056	0.03044	0.2192	0.2625	3.704×10^{-5}	0.7514	0.522	continuation	4.312×10^{-5}	0.064	continuation [§]
0.4328	4.826×10^{-3}	6.190×10^{-3}	0.04252	0.08939	6.457×10^{-9}	0.2454	0.645	continuation	6.565×10^{-4}	0.091	continuation [§]
0.4811	1.871×10^{-4}	5.920×10^{-4}	6.227×10^{-3}	0.01390	4.087×10^{-10}	0.03547	0.818	efficacy stop	1.231×10^{-3}	0.149	efficacy stop
0.5273	1.771×10^{-6}	2.624×10^{-11}	5.820×10^{-10}	3.870×10^{-9}	4.772×10^{-13}	2.453×10^{-9}	0.977	efficacy stop	2.024×10^{-3}	0.237	efficacy stop
0.5504	4.931×10^{-7}	8.238×10^{-8}	2.138×10^{-5}	9.584×10^{-5}	8.399×10^{-15}	7.730×10^{-5}	0.933	efficacy stop	2.518×10^{-3}	0.262	efficacy stop
0.6639	5.942×10^{-8}	6.323×10^{-13}	5.317×10^{-7}	4.641×10^{-6}	4.060×10^{-17}	1.562×10^{-7}	0.960	efficacy stop	5.943×10^{-3}	0.466	efficacy stop
0.7773	3.226×10^{-10}	8.071×10^{-13}	1.643×10^{-12}	2.175×10^{-12}	1.029×10^{-13}	7.248×10^{-12}	0.975	efficacy stop	0.01101	0.703	efficacy stop
0.9013	2.973×10^{-9}	5.968×10^{-10}	1.868×10^{-9}	5.154×10^{-9}	2.368×10^{-11}	8.457×10^{-9}	0.719	efficacy stop	0.01823	0.799	efficacy stop
0.9349	4.081×10^{-9}	2.097×10^{-9}	3.426×10^{-9}	3.991×10^{-8}	9.335×10^{-10}	1.099×10^{-8}	0.354	efficacy stop	0.02044	0.896	efficacy stop
1.0000	2.101×10^{-11}	2.101×10^{-11}	2.101×10^{-11}	0.0000	2.101×10^{-11}	2.101×10^{-11}	0.484	efficacy stop	0.25000	0.941	efficacy stop
1.0000	1.708×10^{-10}	1.836×10^{-11}	1.836×10^{-11}	0.0000	1.836×10^{-11}	1.836×10^{-11}	0.423	efficacy stop	0.25000	0.949	efficacy stop

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 7. Result for OS of trial No.3 [22] with the actual final p-value of log-rank test was 0.34293. From the information fraction of 0.8226, the 95th percentiles of the final predicted values are less than 0.05. The judgement was efficacy stop at the information fraction of 0.8226. This trial would have been suggested to continue until the information fraction reached 0.8547 by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. The judgement of our method and of the conventional method were similar. The values of power were larger than the values of the conditional power when information fraction was less than 1.00 ($P=3.76 \times 10^{-5}$ by t-test).

Information fraction	P-value of log-rank test at the information fraction	Predicted p-value of log-rank test at end of the clinical trial					Power	Judgement by our method	Conventional method		
		Median	Mean	Standard deviation	5 th percentile	95 th percentile			P-value [†]	Conditional power	Judgement [‡]
0.7735	0.0476	0.0428	0.0479	0.0240	0.0073	0.0832	0.466	continuation	0.0108	0.056	continuation [§]
0.8013	0.0888	0.0756	0.0785	0.0331	0.0158	0.1778	0.377	continuation	0.0123	0.047	continuation [§]
0.8226	0.0166	0.0164	0.0233	0.0133	0.0058	0.0449	0.560	efficacy stop	0.0135	0.075	continuation [§]
0.8547	0.0068	0.0053	0.0071	0.0045	7.609×10^{-4}	0.0182	0.680	efficacy stop	0.0153	0.088	efficacy stop
0.8718	0.0034	0.0029	0.0034	0.0016	7.660×10^{-4}	0.0065	0.758	efficacy stop	0.0164	0.100	efficacy stop
0.8932	0.0159	0.0105	0.0187	0.0048	0.0054	0.0239	0.569	efficacy stop	0.0177	0.085	efficacy stop
0.9316	0.0094	0.0149	0.0149	0.0040	0.0046	0.0222	0.608	efficacy stop	0.0202	0.101	efficacy stop
1.0000	0.0057	0.0045	0.0045	0.0000	0.0045	0.0045	0.466	efficacy stop	0.0250	0.018	efficacy stop
1.0000	0.3429	0.3429	0.3429	0.0000	0.3429	0.3429	0.624	futility stop	0.0250	0.136	continuation [§]

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 8. Result for PFS of trial No.4 [23] with the actual final p-value of log-rank test was 3.2349×10^{-12} . From the information fraction of 0.2528, the 95th percentiles of the final predicted values are less than 0.05. The judgement was efficacy stop correctly after the information fraction of 0.213. This trial would have been suggested to stop until the information fraction reached 0.4417 by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. Our method was able to make the right decision as efficacy stop earlier than the conventional method. The values of power were larger than the values of the conditional power ($P=8.75 \times 10^{-6}$ by t-test).

Information fraction	P-value of log-rank test at the information fraction	Predicted p-value of log-rank test at end of the clinical trial					Power	Judgement by our method	Conventional method		
		Median	Mean	Standard deviation	5 th percentile	95 th percentile			P-value [†]	Conditional power	Judgement [‡]
0.2130	0.07933	0.09299	0.1591	0.2125	2.135×10^{-6}	0.6595	0.526	continuation	1.257×10^{-6}	0.046	continuation [§]
0.2528	0.01342	5.956×10^{-6}	4.960×10^{-3}	0.01505	4.585×10^{-8}	0.02623	0.840	efficacy stop	8.269×10^{-5}	0.068	continuation [§]
0.3111	5.353×10^{-3}	3.812×10^{-6}	1.708×10^{-5}	3.382×10^{-5}	1.184×10^{-7}	9.627×10^{-5}	0.940	efficacy stop	5.857×10^{-5}	0.095	continuation [§]
0.3750	1.489×10^{-3}	2.483×10^{-7}	3.258×10^{-6}	1.081×10^{-5}	2.148×10^{-9}	1.563×10^{-5}	0.973	efficacy stop	2.520×10^{-4}	0.127	continuation [§]
0.4417	2.924×10^{-5}	5.818×10^{-7}	4.348×10^{-6}	8.266×10^{-6}	1.206×10^{-8}	2.196×10^{-4}	0.978	efficacy stop	7.445×10^{-4}	0.185	efficacy stop
0.4861	1.961×10^{-5}	3.479×10^{-7}	1.095×10^{-6}	2.130×10^{-6}	4.547×10^{-14}	5.144×10^{-6}	0.978	efficacy stop	1.305×10^{-3}	0.251	efficacy stop
0.5306	0.01803	1.392×10^{-10}	1.169×10^{-6}	4.443×10^{-6}	5.009×10^{-15}	9.464×10^{-6}	0.966	efficacy stop	2.090×10^{-3}	0.318	futility stop
0.5806	4.515×10^{-6}	2.141×10^{-10}	1.526×10^{-8}	4.900×10^{-8}	3.632×10^{-15}	9.921×10^{-8}	0.962	efficacy stop	3.264×10^{-3}	0.385	efficacy stop
0.6778	4.462×10^{-7}	1.445×10^{-9}	6.333×10^{-8}	1.658×10^{-7}	6.247×10^{-12}	4.075×10^{-7}	0.899	efficacy stop	6.478×10^{-3}	0.506	efficacy stop
0.750	7.694×10^{-7}	4.613×10^{-8}	4.613×10^{-7}	7.371×10^{-7}	4.067×10^{-10}	2.284×10^{-6}	0.868	efficacy stop	9.649×10^{-3}	0.551	efficacy stop
0.8333	5.226×10^{-7}	2.173×10^{-6}	1.617×10^{-5}	3.767×10^{-5}	5.443×10^{-9}	7.856×10^{-5}	0.629	efficacy stop	0.01408	0.592	efficacy stop
0.900	3.102×10^{-8}	1.086×10^{-8}	2.457×10^{-7}	4.038×10^{-7}	1.826×10^{-9}	1.186×10^{-6}	0.635	efficacy stop	0.01815	0.695	efficacy stop
0.969	3.221×10^{-10}	1.787×10^{-10}	2.113×10^{-10}	1.215×10^{-10}	9.172×10^{-11}	4.400×10^{-10}	0.610	efficacy stop	0.0282	0.808	efficacy stop

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡: Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence.

Table 9. Result for OS of trial No.4 [24] with the actual final p-value of log-rank test was 0.22017. The 5th and 95th percentiles of the final predicted values by our method straddled the top and bottom across 0.05. The judgement was futility stop at 0.4278 and 0.9250 of information fractions. This trial would have been suggested to continue or stop by futility without statistical confidence at the any stage by using the Lan-DeMets spending function to approximate an O'Brien-Fleming bound. Our method was not able to make the right decision that was futility stop earlier than the conventional method. The values of power were larger than the values of the conditional power (P=0.0002 by t-test).

Information fraction	P-value of log-rank test at the information fraction	Predicted p-value of log-rank test at end of the clinical trial					Power	Judgement by our method	Conventional method		
		Median	Mean	Standard deviation	5 th percentile	95 th percentile			P-value [†]	Conditional power	Judgement [‡]
0.2500	0.2389	1.046×10 ⁻⁴	1.939×10 ⁻⁴	2.467×10 ⁻⁴	1.285×10 ⁻⁵	8.371×10 ⁻⁴	0.931	efficacy stop	7.367×10 ⁻⁶	0.034	continuation [§]
0.4278	0.1477	0.1484	0.1553	0.04611	0.08651	0.2453	0.368	futility stop	6.103×10 ⁻⁴	0.037	continuation [§]
0.5306	0.0180	7.165×10 ⁻³	7.761×10 ⁻³	2.836×10 ⁻³	3.987×10 ⁻³	0.01296	0.682	efficacy stop	2.090×10 ⁻⁴	0.073	continuation [§]
0.5778	0.0107	2.471×10 ⁻³	2.650×10 ⁻³	8.253×10 ⁻⁴	1.517×10 ⁻³	4.253×10 ⁻³	0.803	efficacy stop	3.191×10 ⁻³	0.080	continuation [§]
0.6111	0.0383	1.616×10 ⁻³	1.669×10 ⁻³	4.102×10 ⁻⁴	1.044×10 ⁻³	2.376×10 ⁻³	0.817	efficacy stop	4.141×10 ⁻³	0.065	continuation [§]
0.6694	0.0144	0.0970	0.02036	3.870×10 ⁻³	0.01458	0.02650	0.616	efficacy stop	6.154×10 ⁻³	0.086	continuation [§]
0.7167	0.0168	2.842×10 ⁻³	2.962×10 ⁻³	5.360×10 ⁻⁴	2.308×10 ⁻³	3.951×10 ⁻³	0.749	efficacy stop	8.105×10 ⁻³	0.087	continuation [§]
0.8278	0.0905	0.05458	0.05437	9.701×10 ⁻³	0.03798	0.06900	0.375	continuation	0.01376	0.051	continuation [§]
0.9250	0.0830	0.05975	0.06048	6.886×10 ⁻³	0.05016	0.07521	0.297	futility stop	0.01978	0.054	continuation [§]
0.9917	0.1901	0.2059	0.2051	0.01341	0.1806	0.2259	0.103	continuation [§]	0.02440	0.032	continuation [§]

†; P-value of the Lan-DeMets spending function to approximate an O'Brien-Fleming bound, ‡; Judgement by the Lan-DeMets spending function, §; continuation or futility stop without statistical confidence

FIGURE LEGENDS

Figure 1. Analyzed graphs for PFS of trial No.1 [20]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio gradually increased until the information fraction reached 0.4, and then it remained at almost the final ratio 0.4. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio gradually increased until the information fraction reached 0.4, and then it remained at almost the final ratio 0.6. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 9.351×10⁻¹². Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a

function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). From the early stage of the information fraction of 0.3256, the final predicted value by our method was expected to be below 0.05 and correct.

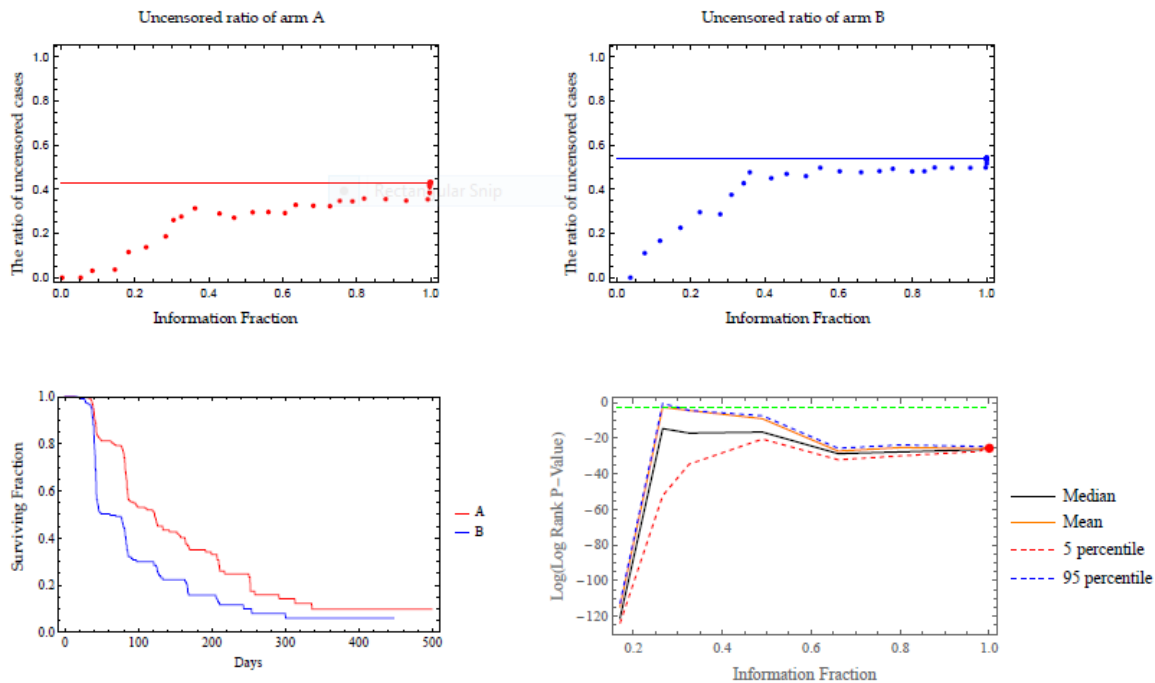


Fig-1

Figure 2. Analyzed graphs for OS of trial No.1 [20]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing until the information fraction reached 1.0. The final value was about 0.2, which is a low value, comparing with other clinical trials. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing until the information fraction reached 1.0. The final value was about 0.2, which is a low value, comparing with other clinical trials. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 0.41185. Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). The log-rank test could not be performed because few events occurred when the information fraction was less than 0.4245. From the early stage of the information, the final predicted value by our method was expected to be over 0.05 and was correct.

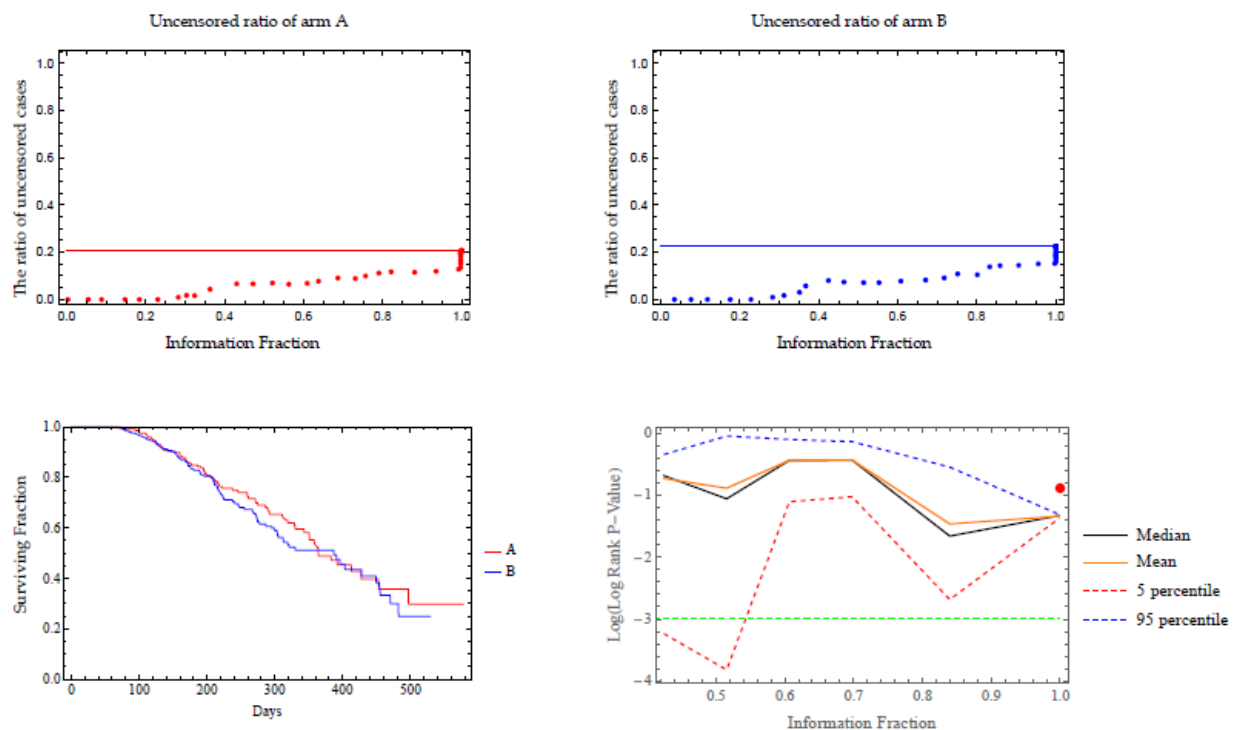


Fig-2

Figure 3. Analyzed graphs for PFS of trial No.2 [21]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing until the information fraction reached 1.0. The final value was about 1.0, which is a high value. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing until the information fraction reached 1.0. The final value was about 1.0, which is a high value. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 0.00194. Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). From the early stage of the information fraction of 0.1733, the final predicted value by our method was expected to be below 0.05 and was correct.

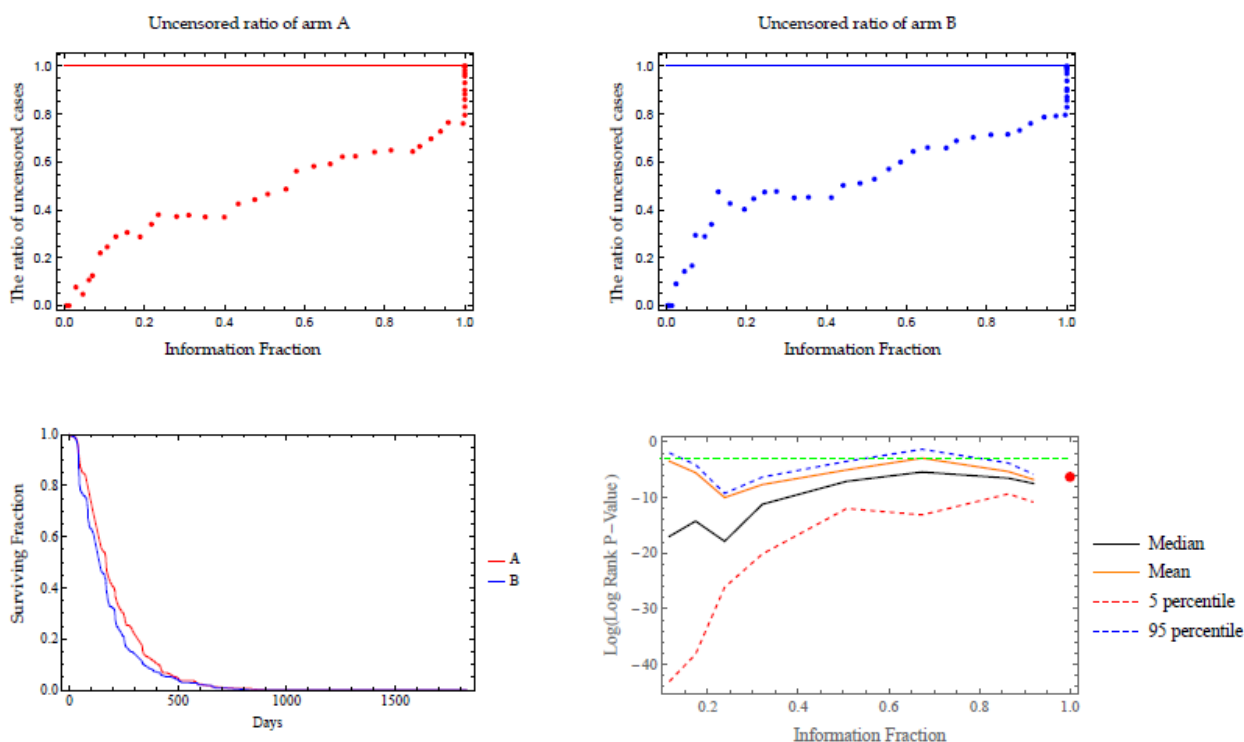
**Fig-3**

Figure 4. Analyzed graphs for OS of trial No.2 [21]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing to about 0.4 until the information fraction reached 1.0, then increased to 1.0. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing to about 0.4 until the information fraction reached 1.0, then increased to 1.0. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 0.41185. Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). The 95th percentiles, median and mean of the final predicted values were greater than or less than 0.5 and the 5th percentiles of the values were less than 0.05. The 95th percentile of the final predicted value by our method was less than 0.05 only at information fraction at 0.432.

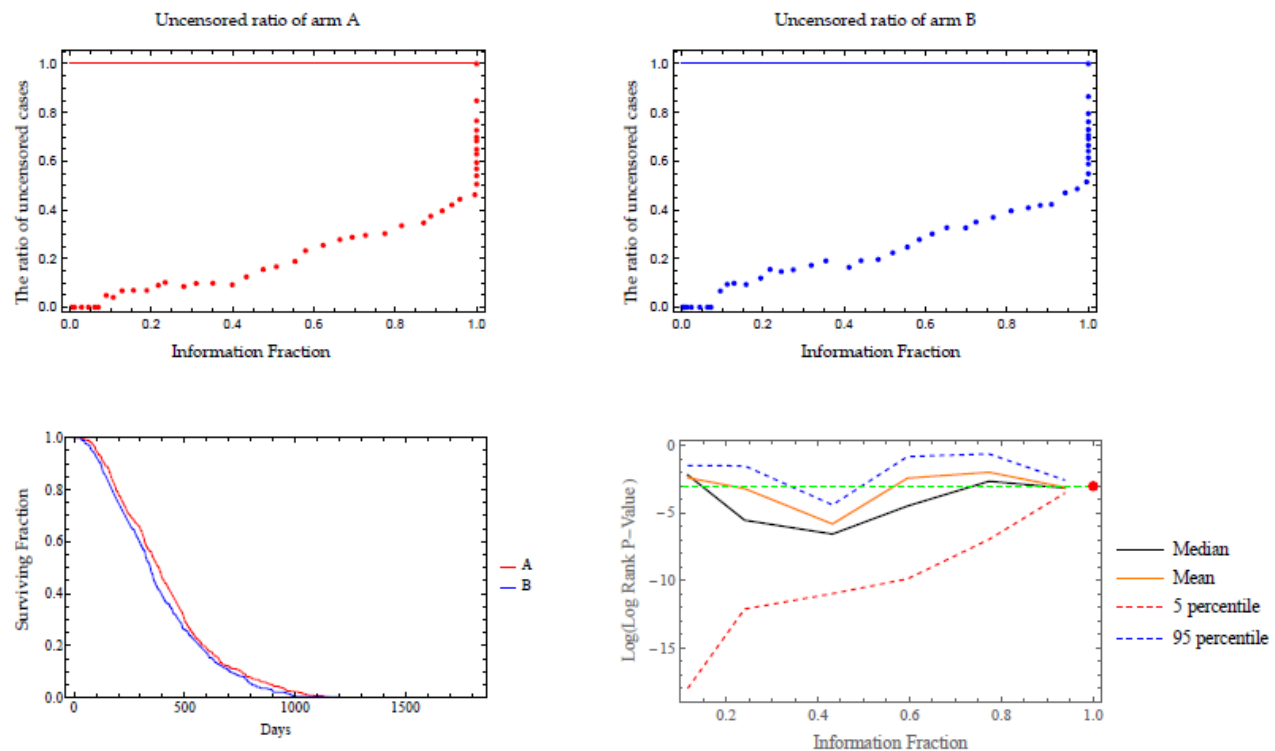


Fig-4

Figure 5. Analyzed graphs for PFS of trial No.3 [22]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing until the information fraction reached 1.0. The final value was about 0.75, which is a high value, comparing with other clinical trials. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing until the information fraction reached 1.0. The final value was about 0.9, which is a high value, comparing with other clinical trials. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 1.8357×10^{-11} . Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). From the information fraction of 0.4811, the final predicted value by our method was expected to be below 0.05 and correct.

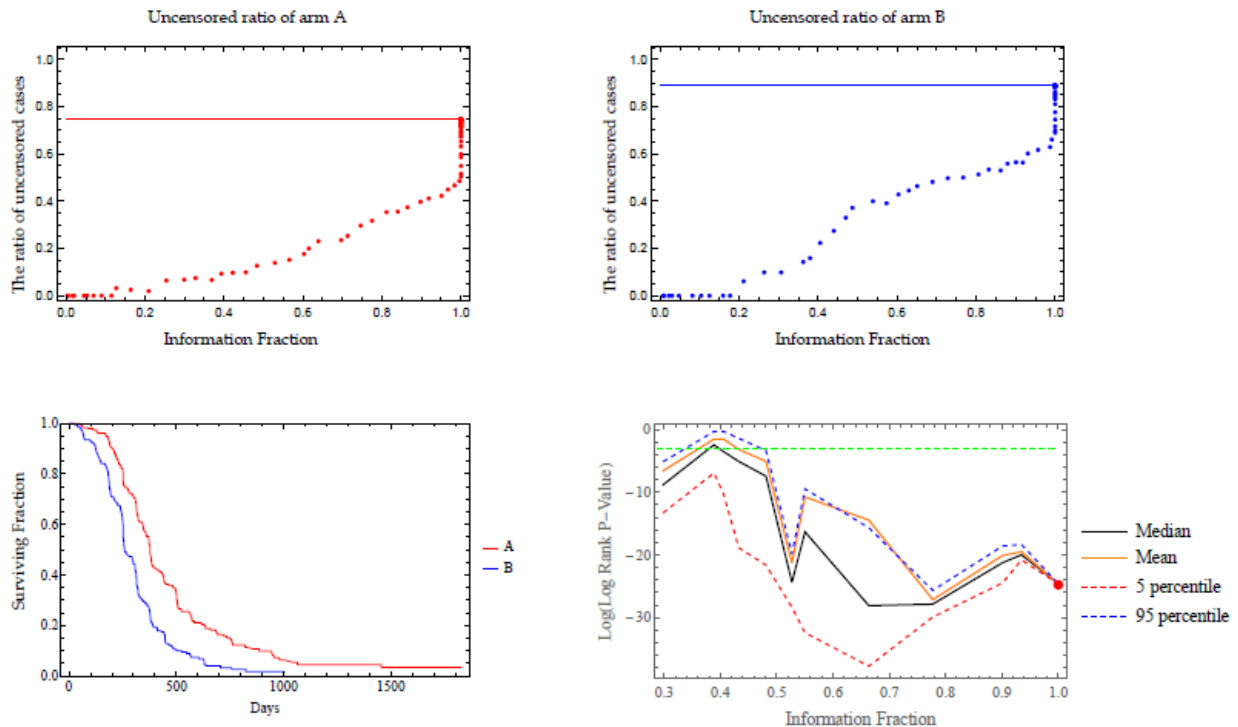
**Fig-5**

Figure 6. Analyzed graphs for OS of trial No.3 [22]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing to about 0.1, which is a low value, until the information fraction reached 1.0. The final value increased to about 0.7. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing to about 0.2, which is a low value, until the information fraction reached 1.0. The final value increased to about 0.7. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 0.34293. Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). The log-rank test could not be performed until the information fraction was less than 0.7735 because few events occurred. From the information fraction of 0.8226 to close to 1.0, the final predicted value by our method was expected to be below 0.05 and was incorrect.

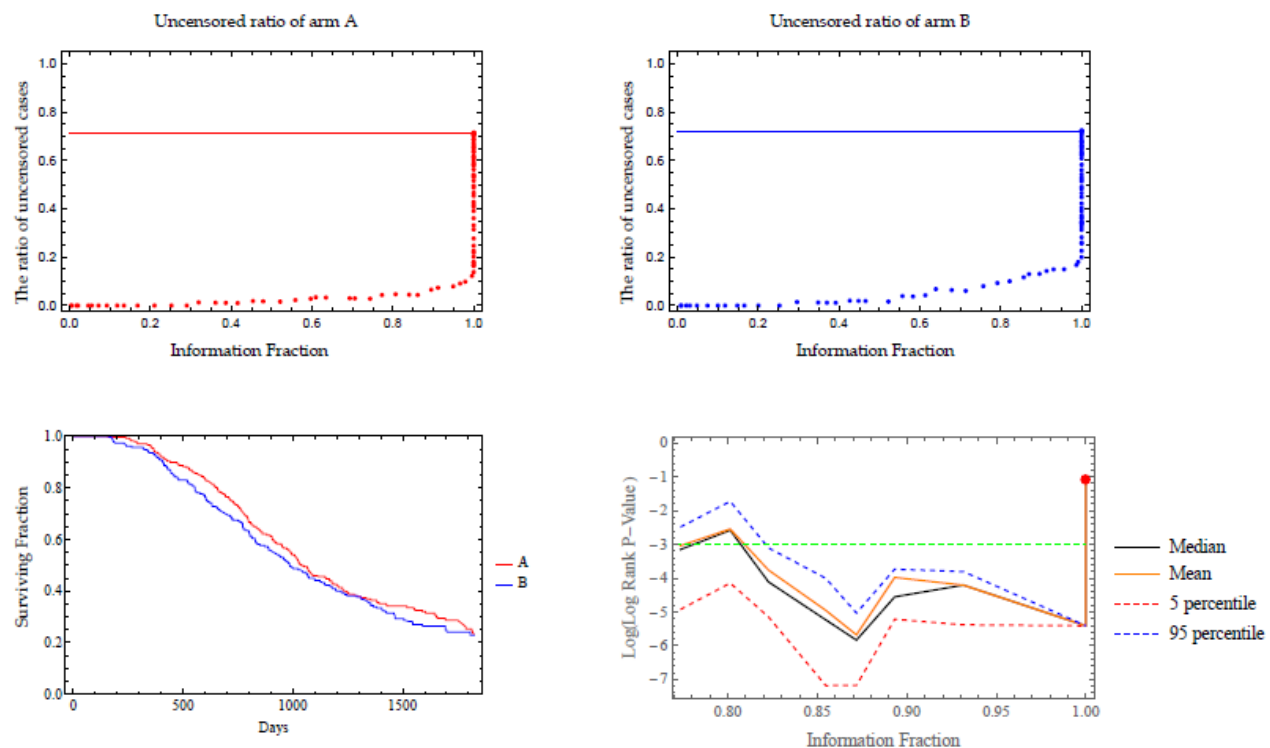


Fig-6

Figure 7. Analyzed graphs for PFS of trial No.4 [23]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio gradually increased until the information fraction reached 0.4, and then it constantly increasing to about 0.7 until the information fraction reached 1.0. The final value increased to 1.0. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio was constantly increasing to about 0.6 until the information fraction reached 1.0. The final value increased to 1.0. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 3.2349×10^{-12} . Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). From the information fraction of 0.2528, the final predicted value by our method was expected to be below 0.05 and correct.

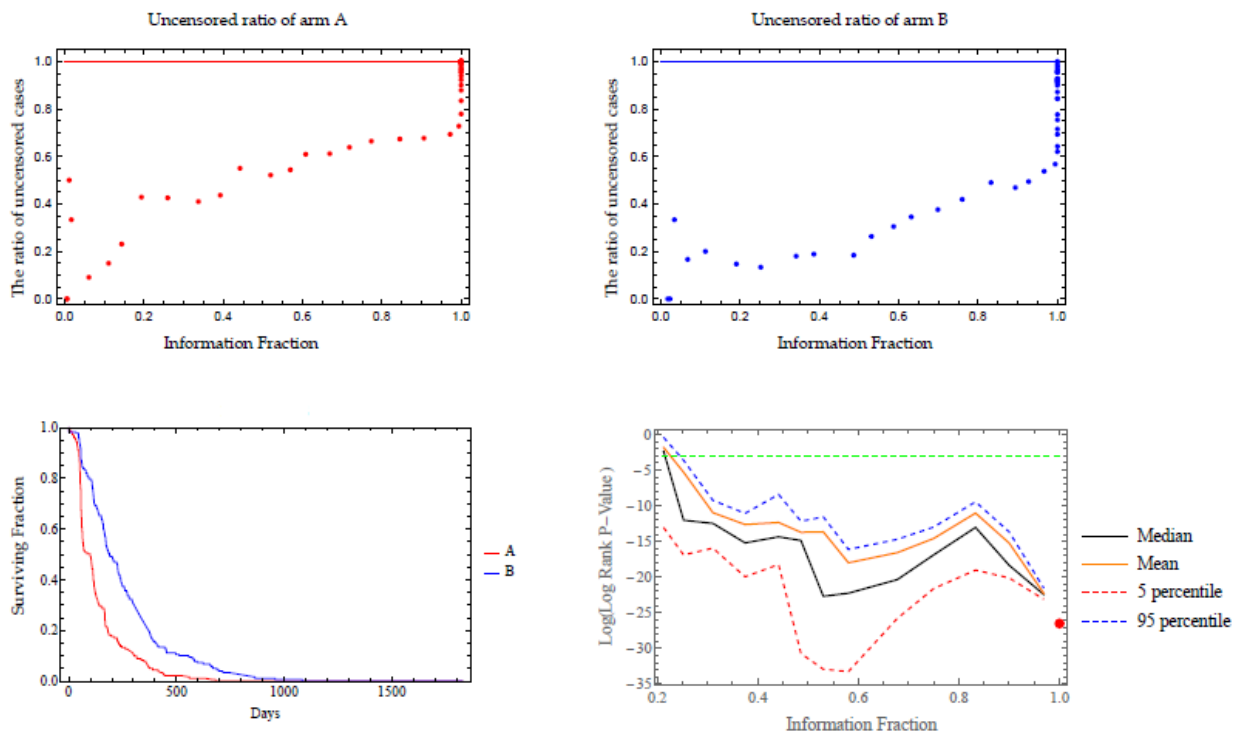
**Fig-7**

Figure 8. Analyzed graphs for OS of trial No.4 [24]. The data was processed at the date of analysis so that the longest observation period does not exceed the observation period; 1826.25 days. Upper left; the ratio of uncensored cases of arm A as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio gradually increased to about 0.2 until the information fraction reached 1.0. The final value increased to about 0.7. Upper right; the ratio of uncensored cases of arm B as a function of information fraction. The interval between the points is equivalent to one month. The horizontal bar indicates the final ratio when the study was finished. The ratio gradually increased to about 0.2 until the information fraction reached 1.0. The final value increased to about 0.8. Lower left; Kaplan-Meier survival curve at the end of the clinical trial (Arm A; red, arm B; blue). The p-value of log-rank test was 0.22017. Lower right; The logarithm of p-values of predicted log-rank tests at the end of the clinical trial as a function of information fraction (median; black, mean; red, 5th percentile; dashed red, 95th percentile; dashed blue, 0.05 that is the boundary for significance; dashed green, final p-value of log-rank test; red circle). The 95th percentiles, median and mean of the final predicted values value by our method was greater than or less than 0.05. The 95th percentiles of the final predicted values were less than 0.05 between 0.5306 and 0.7167 of information fraction.

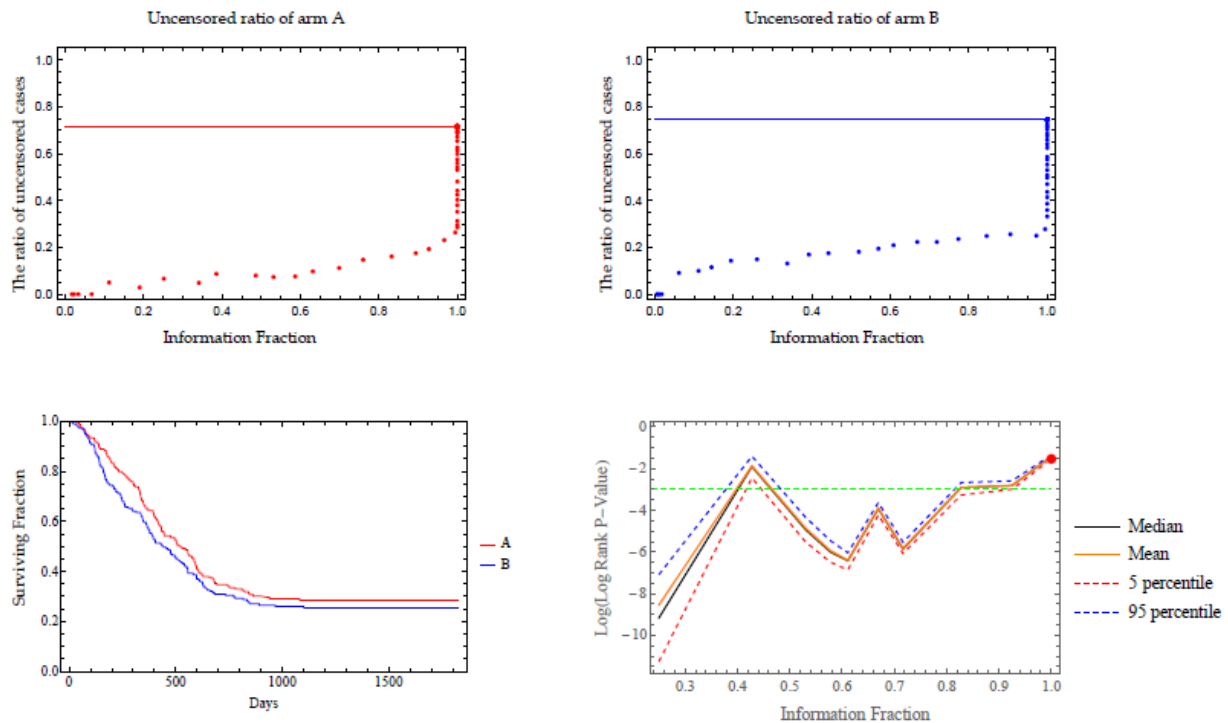


Fig-8

This study was supported by no funds. This study was not performed a part of the employment of the authors. There was no funder in this study.

References

- [1]. Simon, R. (1989) Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*, **10**, 1-10.
- [2]. Jung, S. H., Lee, T., Kim, K. and George, S. L. (2004). Admissible two-stage designs for phase II cancer clinical trials. *Statist in Med.*, **23**, 561-569.
- [3]. Aspril, J. (2018) Clinical trials to obtain FDA approval typically account for small proportion of total drug research and development costs, study suggests. *Johns Hopkins Bloomberg School of Public Health*, <https://publichealth.jhu.edu/2018/cost-of-clinical-trials-for-new-drug-FDA-approval-are-fraction-of-total-tab>
- [4]. DiMasi, J. A., Grabowski, H. G. and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Economics*, **47**, 20-33., <https://doi.org/10.1016/j.jhealeco.2016.01.012>.
- [5]. Gordon Lan, K. K. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, **70**, 659-663.
- [6]. Pocock, S.J. (1977) Group sequential method in the design and analysis of clinical trials. *Biometrika*, **64**, 191-199.
- [7]. O'Brien, P.C. & Fleming, T.R. (1979) A multiple testing procedure for clinical trials. *Biometrics*, **35**, 549-612.

- [8]. Uemura, K., Matsuyama, Y. & Ohashi, Y. (2008) A Modification of the 50%-Conditional Power Approach for Increasing the Sample Size Based on an Interim Estimate of Treatment Difference. *Jpn J Biometrics*, **29**, 19-34.
- [9]. Jitlal, M., Khan, I., Lee, S. M. and Hackshaw, A. (2012). Stopping clinical trials early for futility: retrospective analysis of several randomised clinical studies. *Br J Cancer*, **107**, 910-917.
<https://doi.org/10.1038/bjc.2012.344>
- [10]. Chen, Y.H., DeMets, D.L. and Gordon Lan, K.K. (2004) Increasing sample size when the unblinded interim result is promising. *Statist in Medicine*, **23**, 1023-1038.
- [11]. Candès, E.J., Romberg, J. and Toa, T. (2006) Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory*, **52**, 489-509.
- [12]. Donoho, D.L. (2006) Compressed sensing. *IEEE Trans. Inf. Theory*, **52**, 1289-1306.
- [13]. Candès, E.J. and Tao, T. (2006) Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, **52**, 5406-5425.
- [14]. Rani, M., Dhok, S.B. & Deshmukh, R.B. (2018) A systematic review of compressive sensing: Concepts, implementations and applications. *IEEE Access*, **6**, 4875-4894.
- [15]. Miyagi, Y., Fujiwara, K., Oda, T., Miyake, T. and Coleman, R. L. (2018). Development of new method for the prediction of clinical trial results using compressive sensing of artificial intelligence. *J Biostat Biometric.*, **3**, 202.
- [16]. Miyagi, Y., Fujiwara, K., Oda, T., Miyake, T. and Coleman, R.L. (2020) Studies on Development of New Method for the Prediction of Clinical Trial Results Using Compressive Sensing of Artificial Intelligence. In: *Theory and Practice of Mathematics and Computer Science* (ed M.A.M. Ferreira), pp101-108. Hooghly, West Bengal, India: Book Publisher International. Print ISBN: 978-93-90206-63-6, eBook ISBN: 978-93-90206-64-3 DOI: 10.9734/bpi/tpmcs/v2 <https://bp.bookpi.org/index.php/bpi/catalog/book/271>
- [17]. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A., et al. (2017) Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Annals of Internal medicine*. **167**, 63-5.
- [18]. Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G. and Tierney, J. F. (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *Jama.*, **313**, 1657-1665. DOI: 10.1001/jama.2015.3656.
- [19]. Hussain, J. A., White, I. R., Johnson, M. J., Bland, M. and Currow, D. C. (2021). Performance status and trial site-level factors are associated with missing data in palliative care trials: An individual participant-level data analysis of 10 phase 3 trials. *Palliative medicine*, **35**, 1998-2007. doi: 10.1177/02692163211040970. Epub 2021 Sep 16.
- [20]. Van Cutsem, E., Yoshino, T., Lenz, H.J., Lonardi, S., Falcone, A., Limón, M.L., et al. (2018) Nintedanib for the treatment of patients with refractory metastatic colorectal cancer (LUME-Colon 1): a phase III, international, randomized, placebo-controlled study. *Ann Oncol.*, **29**, 1955-1963.
<https://doi.org/10.1093/annonc/mdy241>.
- [21]. Tabernero, J., Yoshino, T., Cohn, A. L., Obermannova, R., Bodoky, G., Garcia-Carbonero, R., et al. (2015). Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): a randomised, double-blind, multicentre, phase 3 study. *Lancet Oncol.*, **16**, 499-508.

- [22]. Aghajanian, C., Goff, B., Nycum, L.R., Wang, Y.V., Husain, A. and Blank, S.V. (2015) Final overall survival and safety analysis of OCEANS, a phase 3 trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent ovarian cancer. *Gynecol. Oncol.* **139**, 10-16. doi: 10.1016/j.ygyno.2015.08.004.
- [23]. Pujade-Lauraine, E., Hilpert, F., Weber, B., Reuss, A., Poveda, A., Kristensen, G., et al. (2014). Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: the AURELIA open-label randomized phase III trial. *Obstet Gynecol Survey*, **69**, 402-404.
- [24]. Bamias, A., Gibbs, E., Khoon Lee, C., Davies, L., Dimopoulos, M., Zagouri, F., et al. (2017) Bevacizumab with or after chemotherapy for platinum-resistant recurrent ovarian cancer: exploratory analyses of the AURELIA trial. *Ann. Oncol.*, **28**, 1842-1848. doi: 10.1093/annonc/mdx228.
- [25]. Bennouna, J., Sastre, J., Arnold, D., Österlund, P., Greil, R., Van Cutsem, E., et al. (2013) ML18147 Study Investigators. Continuation of bevacizumab after first progression in metastatic colorectal cancer (ML18147): a randomised phase 3 trial. *Lancet Oncol.* **14**, 29-37. doi: 10.1016/S1470-2045(12)70477-1.
- [26]. Miyagi, Y., Tada, K., Yasuhi, I., Maekawa, Y., Okura, N., Kawakami, K., et al. (2020) New method for determining fibrinogen and FDP threshold criteria by artificial intelligence in cases of massive hemorrhage during delivery. *J Obstet Gynaecol Res.*, **46**, 256–265. <https://doi.org/10.1111/jog.14166>
- [27]. Forbes, C., Evans, M., Hastings, N. & Peacock, B. (2011). Statistical distributions (Vol. 4). New York: Wiley.
- [28]. Chen, M.Z., Chen, H., Ma, Y.C., Wang, B. and Tang, Y. (2011). Recovery rate Models Based on General Beta Regression for Non-performing Loan. *J Applied Stat Management*, **30**, 810–823.
- [29]. Zhai, Z., Jia, G. & Wang, K. (2018) A Novel Teaching-Learning-Based Optimization with Error Correction and Cauchy Distribution for Path Planning of Unmanned Air Vehicle. *Comput Intell Neurosci.*, **2018**, Article ID 5671709. <https://doi.org/10.1155/2018/5671709>
- [30]. Anderson, D. (1981) Maximum Likelihood Estimation in the Non-Central Chi Distribution with Unknown Scale Parameter. *Sankhyā: The Ind. J. Stat. Series B.*, **43**, 58-67.
- [31]. Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, **2**, 110-114.
- [32]. Epstein B. and Sobel M. (1954) Some Theorems Relevant to Life Testing from an Exponential Distribution. *Ann Math Statist.*, **25**, 373-381. doi:10.1214/aoms/1177728793.
- [33]. Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251-261.
- [34]. Nadarajah, S. & Kotz, S. (2003) The exponentiated Fréchet distribution. *Interstat Electronic J.*, **14**, 1-7.
- [35]. Stacy, E.W. (1962) A Generalization of the Gamma Distribution. *Ann Math Statist.*, **33**, 1187–1192. doi:10.1214/aoms/1177704481. <https://projecteuclid.org/euclid.aoms/1177704481>
- [36]. Nadarajah, S. & Kotz, S. (2004) The beta Gumbel distribution. *Math Problems in engineering*, **4**, 323-332.
- [37]. Cooray, K. and Ananda, M. (2008) A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics—Theory and Methods*, **37**, 1323-1337.
- [38]. Chhikara, R.S. and Folks, J.L. (1977) The inverse Gaussian distribution as a lifetime model. *Technometrics*, **19**, 461-468.
- [39]. Barndorff-Nielsen, O.E. (1997) Normal inverse Gaussian distributions and stochastic volatility models, *Scand. J. Statist.*, **24**, 1–13.

-
- [40]. Reed, W.J. (2006) The Normal-Laplace Distribution and Its Relatives. In: *Advances in Distribution Theory, Order Statistics, and Inference. Statistics* (eds N. Balakrishnan, J.M. Sarabia and E. Castillo), pp. 61-74. Boston: Birkhäuser.
- [41]. Coronel-Brizio, H.F. & Hernandez-Montoya, A.R. (2005) On fitting the Pareto-Levy distribution to stock market index data: Selecting a suitable cutoff value. *Physica A: Stat Mech Appl*, **354**, 437-449.
- [42]. Gupta, S. S. and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika*, **53**, 565-570.
- [43]. Mudholkar, G. S. and George, E. O. (1978). A remark on the shape of the logistic distribution. *Biometrika*, **65**, 667-668.
- [44]. Limpert, E., Stahel, W. A. and Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. *BioScience*, **51**, 341-352.
- [45]. Koch AL. (1966). The logarithm in biology. I. Mechanisms generating the log-normal distribution exactly. *J Theoret Biol* **23**, 276-290.
- [46]. Reed, W.J. & Jorgensen, M. (2004) The double Pareto-lognormal distribution: A new parametric model for size distributions. *Communications in Statistics—Theory and Methods*, **33**, 1733-1753.
- [47]. Roy, D. (2004) Discrete rayleigh distribution. *IEEE Transactions on Reliability*. **53**, 255-260.
- [48]. Ord, J.K. (1968) The Discrete Student's *t* Distribution. *The Annals of Mathematical Statistics*, **39**, 1513-1516.
- [49]. Avigad, J. (2013) Uniform distribution and algorithmic randomness. *The J. Symbolic. Logic.*, **78**, 334-344.
- [50]. Miyagi, Y. and Miyake, T. (2020) Potential of artificial intelligence for estimating Japanese fetal weights. *Acta Medica Okayama*, **74**, 483-493.
- [51]. Miyagi, Y., Habara, T., Hirata, R. and Hayashi, N. (2019) Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. *Reprod Med Biol.*, **18**, 190-203.
- [52]. Miyagi, Y., Habara, T., Hirata, R. and Hayashi, N. (2019). Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. *Reprod Med Biol.*, **18**, 204-211.
- [53]. Miyagi, Y., Habara, T., Hirata, R. and Hayashi, N. (2019). Feasibility of predicting live birth by combining conventional embryo evaluation with artificial intelligence applied to a blastocyst image in patients classified by age. *Reprod Med Biol.*, **18**, 344-35.
- [54]. Miyagi, Y., Habara, T., Hirata, R. and Hayashi, N. (2020). Predicting a live birth by artificial intelligence incorporating both the blastocyst image and conventional embryo evaluation parameters. *Artif Intell Med Imaging*, **1**, 94-107.
- [55]. Miyagi, Y., Habara, T., Hirata, R. and Hayashi, N. (2022) New methods for comparing embryo selection methods by applying artificial intelligence: Comparing embryo selection AI for live births. *Br J Healthcare Med Res.*, **9**, 36-44.
- [56]. Miyagi, Y., Habara, T., Hirata, R. and Hayashi, N. (2022). Deep Learning to Predicting Live Births and Aneuploid Miscarriages from Images of Blastocysts Combined with Maternal Age. *Int J Bioinform Intellig Comput.*, **1**, 10-21.
-

- [57]. Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A. and Miyake, T. (2021). Recognition of facial expression of fetuses by artificial intelligence (AI). *J Perinat Med*, **49**, 596-603.
- [58]. Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A. and Miyake, T. (2021). Recognition of fetal facial expressions using artificial intelligence deep learning. *Donald School J Ultrasound Obstet Gynecol* **15**, 223-228.
- [59]. Miyagi, Y., Takehara, K. and Miyake, T. (2019) Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images. *Mol Clin Oncol.*, **11**, 583-589.
- [60]. Miyagi, Y., Takehara, K., Nagayasu, Y. and Miyake, T. (2020). Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. *Oncol Lett.*, **19**, 1602-1610.
- [61]. Miyagi, Y., Hata, T, Bouno., S, Koyanagi. and Miyake, T. (2022) Artificial intelligence to understand fluctuation of fetal brain activity by recognizing facial expressions. *Int J Gynecol Obstet*. doi: 10.1002/ijgo.14569
- [62]. Müller, H.H. and Schäfer, H. (2001) Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*, **57**, 886-891. doi: 10.1111/j.0006-341x.2001.00886.x.
- [63]. Lehman, W. and Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics*, **55**, 1286-1290. doi: 10.1111/j.0006-341x.1999.01286. x. PMID: 11315085.
- [64]. Wassmer, G. (2006) Planning and analyzing adaptive group sequential survival trials. *Biom J.*, **48**, 714-729. doi: 10.1002/bimj.200510190. PMID: 16972724.
- [65]. Bauer, P. and Kohne, K. (1994) Evaluation of Experiments with Adaptive Interim Analyses. *Biometrics*, **50**, 1029–1041. <https://doi.org/10.2307/2533441>
- [66]. Chang, M. (2007) Adaptive design method based on sum of p-values. *Statist in Med.*, **26**, 2772-2784.
- [67]. Chow, S.C. and Chang, M. (2008) Adaptive design methods in clinical trials – a review. *Orphanet J Rare Dis.*, **3**, 11. <https://doi.org/10.1186/1750-1172-3-11>
- [68]. Shun, Z., Lan, K.G. and Soo, Y. (2008) Interim treatment selection using the normal approximation approach in clinical trials. *Statist in Med.*, **27**, 597-618.
- [69]. Jenkins, M., Stone, A. and Jennison, C. (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical statistics*, **10**, 347-356.
- [70]. Bauer, P. and Kieser, M. (1999) Combining different phases in the development of medical treatments within a single trial. *Statist. in Med.*, **18**, 1833-1848.