Paper AS10

Revolutionizing Cancer Trials with Adaptive Designs and Bayesian Analysis

A Bayesian Adaptive Approach for Relapsed/Refractory Acute Myeloid Leukemia (AML)

Navneet Jha, Ephicacy Lifescience Analytics Pvt, Ltd., Bengaluru, India

ABSTRACT

Cancer trials are undergoing a transformation with Bayesian adaptive designs, allowing real-time modifications based on accumulating data. This research explores the application of Bayesian adaptive designs in cancer trials, focusing on relapsed/refractory Acute Myeloid Leukemia (AML). Using Kaplan-Meier survival analysis, Bayesian Cox Proportional Hazards models, and Bayesian predictive probability, we evaluate overall survival (OS) and progression-free survival (PFS) as key endpoints.

The results indicate that while Kaplan-Meier survival curves show modest improvements in OS and PFS for the experimental treatment, Bayesian predictive probability estimates (~99%) strongly support trial continuation. This highlights the potential of Bayesian adaptive designs to optimize trial efficiency by enabling real-time decision-making. Unlike traditional methods such as Kaplan-Meier survival analysis, which provides retrospective survival probabilities, Bayesian adaptive methods allow continuous updating of predictive probabilities. This enables early stopping for inefficacy, patient reallocation, and trial adjustments—reducing patient exposure to ineffective treatments and accelerating drug development.

INTRODUCTION

Cancer trials often face challenges such as long durations, high costs, and inefficiencies in patient recruitment. Traditional fixed designs fail to adapt to emerging data, leading to delays in identifying effective treatments. Bayesian adaptive designs offer a solution by allowing real-time modifications based on accumulating data, enabling early stopping for inefficacy, sample size adjustments, and reallocation of patients to promising treatment arms.

This research applies Bayesian Adaptive Trial Design to an AML trial, focusing on OS and PFS as primary endpoints. By integrating Kaplan-Meier survival analysis, Bayesian Cox Proportional Hazards models, and Bayesian predictive probability, we aim to determine whether the trial should continue, expand, or stop early. The research aligns with FDA and EMA guidelines, demonstrating the potential of Bayesian adaptive designs to revolutionize cancer trials.

THE ROLE OF BAYESIAN ADAPTIVE DESIGNS

Bayesian adaptive trials offer a flexible framework for real-time decision-making, improving trial efficiency and patient safety. The key advantages include:

Early Stopping for Inefficacy

- In traditional trials, treatments continue even if they are not working
- Bayesian adaptive trials check the data as it comes in. If treatment is unlikely to work (e.g., success probability falls below 50%), the trial can stop early. This prevents patients from getting ineffective treatments.

Sample Size Adjustments

- Fixed trials set the number of patients at the start, which can be too many or too few.
- Bayesian methods adjust the sample size based on ongoing results. If more data is needed, the trial can add patients. If enough evidence is gathered early, the trials can be continued and completed without adding more.

Reallocation of Patients to Promising Arms

- In regular trials, patients are randomly assigned to different treatments.
- Bayesian trials update treatment effectiveness in real time. If one treatment is working better, more patients can be given that treatment instead of the less effective one. This helps more people get the best possible care

RESEARCH OBJECTIVES

This research applies Bayesian adaptive modeling to an AML trial, using:

- Kaplan-Meier survival estimates for OS/PFS
- Bayesian Cox PH models for hazard ratios
- Bayesian predictive probability for early stopping decisions

Key Question: Should the trial continue based on Bayesian adaptive criteria?

METHOD

REASERCH POPULATION AND DATA

To facilitate the Bayesian Adaptive Trial Analysis for Acute Myeloid Leukemia (AML), we generated a synthetic AML ADTTE (Analysis Dataset for Time-to-Event Endpoints) dataset. This dataset simulates a clinical trial evaluating Progression-Free Survival (PFS) and Overall Survival (OS) using Bayesian adaptive designs. The dataset was developed following statistical simulation principles, ensuring balanced treatment groups, realistic patient characteristics, and adherence to real-world survival data distributions. The dataset was designed to reflect real-world AML clinical trial conditions while incorporating Bayesian Adaptive Design principles. You can find SAS code in the appendix section and data availability: The dataset used in this study is available for access via Google Drive. Researchers and reviewers can request access at [https://docs.google.com/spreadsheets/d/1OmICcSPdy6BlV-A-WsUWl51IhtM7CueZPfZReJA03H8/edit?usp=sharing]

This research has analyzed data from a randomized clinical trial involving patients diagnosed with relapsed or refractory acute myeloid leukemia (AML). The randomized clinical trial contained 100 patients, 44 in the control group and 56 in the experimental treatment group. Patients were monitored for overall survival (OS) and progression-free survival (PFS), with censoring applied where necessary. Total Patients: 100, Control Group: 44 patients, Experimental Group: 56 patients, Median Follow-up Time: 627 days.

ENDPOINTS:

- Overall Survival (OS): Time from treatment to death.
- Progression-Free Survival (PFS): Time to disease progression or death.

CENSORING CRITERIA:

- 0 = Event occurred
- 1 = Censored.

CENSORING RATE:

- Control Group: 11.36%
- Experimental Group: 25.00%

Censoring occurs when patients do not experience the event within the research period, either due to loss to follow-up or remaining event-free at the last recorded visit. The higher censoring rate in the experimental group suggests a potential survival benefit but requires further validation.

Baseline characteristics, including age, prior treatments, disease severity, WBC count, HGB levels, and cytogenetics risk, were comparable between the control and experimental groups

BASELINE CHARACTERISTICS SUMMARY

- Age (years): Control: 57.19 (±9.96), Experimental: 56.94 (±8.87): ensures that age-related effects are balanced
- Prior Treatment (%): Control: 80.7%, Experimental: 70.5%: A slightly higher percentage of prior treatment in the control group, but overall, both groups had substantial prior exposure.
- Disease Severity (%): Moderate and severe cases were equally distributed, confirming no major imbalance.
    - Control: Moderate (45.5%), Severe (54.5%).
    - Experimental: Moderate (42.0%), Severe (58.0%).
- Baseline WBC Count (10^9/L): Control: 50.30 (±28.88), Experimental: 47.83 (±29.53).
- Baseline HB (g/dL): Control: 12.18 (±3.53), Experimental: 11.95 (±3.51).
- Cytogenetics Risk (%): Cytogenetics risk refers to chromosomal abnormalities associated with AML prognosis, which may influence treatment response and survival outcomes. the distribution of Low, Intermediate, and High risk is similar across both groups, ensuring genetic risk factors are balanced as below
    - Control: Low (25.0%), Intermediate (48.9%), High (26.1%).
    - Experimental: Low (30.4%), Intermediate (44.6%), High (25.0%).

These results indicate no significant baseline differences, meaning any differences in survival or treatment response can be attributed to the treatment itself rather than pre-existing imbalances.

STATISTICAL METHODS:

KAPLAN MEIER SURVIVAL ANALYSIS:

The Kaplan-Meier method was used to estimate survival probabilities for OS and PFS. The following SAS code was used:

```
/* Kaplan-Meier Survival Analysis for OS/PFS */
PROC LIFETEST DATA=aml_adtte PLOTS=SURVIVAL;
    TIME AVAL * CNSR(1);  /* 0 = event, 1 = censored */
    STRATA TRT01A;  /* Compare Control vs. Experimental */
    WHERE PARAMCD = "&varname";
RUN;
```

The results obtained from this program can be found in:

- Table 3: Kaplan-Meier Survival Analysis Summary
- Fig 1: Kaplan-Meier Survival Curve for OS
- Fig 2: Kaplan-Meier Survival Curve for PFS

BAYESIAN COX PROPORTIONAL HAZARDS MODEL:

A Bayesian Cox PH model was used to estimate hazard ratios (HR) for the experimental treatment. The following SAS code was used:

```
/* Bayesian Cox Proportional Hazards Model for PFS/OS */
PROC PHREG DATA=aml_adtte;
    CLASS TRT01A;
    MODEL AVAL * CNSR(1) = TRT01A;
    WHERE PARAMCD = "&varname";
    BAYES SEED=12345 NBI=1000 NMC=10000 THIN=10 OUTPOST=Bayesian_Cox;
RUN;
```

Keyword Explanations:

- BAYES: Enables Bayesian estimation instead of conventional maximum likelihood estimation (MLE).
- SEED=12345: Sets a random seed to ensure reproducibility of results.
- NBI=1000 (Number of Burn-in Iterations): The first 1,000 iterations are discarded to allow the Markov Chain to stabilize.
- NMC=10000 (Number of Monte Carlo Iterations): Generates 10,000 posterior samples to estimate parameter distributions.
- OUTPOST=Bayesian_Cox: Stores the posterior distributions and results in a dataset for further analysis.

The results obtained from this program can be found in:

- Table 4: Bayesian Cox PH Model Estimates
- Fig 3: Hazard Ratio Posterior Distributions OS
- Fig 4: Hazard Ratio Posterior Distributions PFS

Feature	Bayesian Cox PH Model	Conventional Cox PH Model
Estimation Method	Bayesian: Uses MCMC sampling to estimate hazard ratios from a posterior distribution.	Frequentist: Uses maximum likelihood estimation (MLE) to derive single-point estimates.
Incorporation of Prior Knowledge	Allows prior distributions on parameters, refining estimates with existing knowledge. E.g. Iincorporate prior knowledge about hazard ratios	Does not incorporate prior information, relying solely on observed data
Uncertainty Measurement	Provides full posterior distributions of hazard ratios with credible intervals (highest posterior density (HPD) intervals).	Reports confidence intervals, but these may be less informative with small sample sizes.
Computation	Uses Markov Chain Monte Carlo (MCMC) simulations	Closed-form solution using likelihood maximization
Interpretation	Provides posterior probabilities, making inference more intuitive	Provides p-values for significance testing
Flexibility in Handling Small Samples	Performs better with small datasets by borrowing information from priors.	Requires larger sample sizes for stable estimates.

Regulatory Acceptance	Increasingly accepted by FDA/EMA but requires justification for priors.	Standard approach in clinical trials and widely accepted.

Table 1: Comparison with Conventional Cox PH Model

BAYESIAN PREDICTIVE PROBABILITY FOR INTERIM DECISIONS:
Bayesian predictive probability was used to determine whether the trial should continue, expand, or stop early. The following SAS code was used:

```
TITLE "Bayesian Predictive Probability for OS/PFS";
PROC MCMC DATA=aml_adtte NBI=1000 NMC=10000 THIN=10 SEED=12345 OUTPOST=Bayes_Predictive_&paramcd;
    WHERE PARAMCD = "&paramcd";
    PARMS theta;
    PRIOR theta ~ BETA(1,1);
    MODEL AVAL ~ NORMAL(theta, VAR=100);
RUN;
```

Keyword Explanations:

- PROC MCMC: Specifies the Bayesian MCMC procedure for parameter estimation.
- DATA=aml_adtte: Defines the dataset containing time-to-event (survival) data.
- NBI=1000: Number of burn-in iterations (first 1,000 samples discarded for convergence).
- NMC=10000: Total number of Monte Carlo simulations (10,000 posterior samples generated).
- THIN=10: Retains every 10th sample to reduce autocorrelation in the MCMC chain.
- SEED=12345: Ensures reproducibility by setting a fixed random seed.
- OUTPOST=Bayes_Predictive: Saves posterior distributions for further analysis.
- PARMS theta: Declares θ (theta) as the parameter to estimate.
- PRIOR theta ~ BETA(1,1): Sets a Beta(1,1) prior, assuming no prior knowledge about the success probability.
- MODEL AVAL ~ NORMAL (theta, VAR=100): Defines the likelihood function where AVAL (time-to-event data) follows a normal distribution centered around theta with a variance of 100.

The results obtained from this program can be found in:

- Table 5: Bayesian Predictive Probability (Θ)
- Fig 5/6: Posterior density plot for OS
- Fig 7/8: Trace plot (iteration vs. Theta) for OS/PFS

This Bayesian predictive model enables real-time probability updates, making it ideal for adaptive trials, where early decisions on trial continuation or stopping can be made based on accumulating data.

Feature	**Frequentist Methods**	**Bayesian Predictive Model**
Inference Type	Point estimates & confidence intervals	Full posterior distribution of treatment success
Flexibility	Fixed sample size	Allow real-time updates as new data arrives
Decision Criteria	Hypothesis testing (p-value, HR)	Predictive probability of success (e.g., **P (θ > 0.85)**)

Table 2: How This Differs from Conventional Approaches?

ADAPTIVE DECISION RULES:

The following rules were applied based on Bayesian probability (θ):

- If θ > 85% → Continue trial.
- If 50% ≤ θ ≤ 85% → Expand sample size.
- If θ < 50% → Stop trial.

The decision rules (θ thresholds) are based on established Bayesian adaptive trial methodologies, as referenced in Berry (2006) and FDA Adaptive Design Guidelines (2019). These thresholds help balance safety, efficacy, and trial efficiency, ensuring ethical and regulatory compliance.

RESULTS:

Treatment Group	Median OS	95% CI (OS)	Median PFS	95% CI (PFS)
Control	596	(520, 702)	377	(249, 438)
Experimental	627	(480, 742)	479	(359, 540)

Table 3: Kaplan-Meier Survival Analysis.

OS = Overall Survival, PFS = Progression-Free Survival, CI = Confidence Interval. All values are in days.

- The **p-value (0.9782)** from the Log-Rank test clearly indicates no significant difference in Overall Survival (OS) between the experimental and control groups.
- The Wilcoxon **test (p = 0.9190)** suggests no significant difference in Progression-Free Survival (PFS) between the experimental and control groups.

The Log-Rank test was used for Overall Survival (OS) as it assumes proportional hazards and is standard for long-term survival comparisons, while the Wilcoxon test was applied to Progression-Free Survival (PFS) because it gives greater weight to early differences, making it more sensitive to early treatment effects. The p-values (OS: 0.9782, PFS: 0.9190) indicate no statistically significant differences between groups; however, the experimental group showed a median PFS improvement of over 100 days (~15%), which may have clinical relevance despite not reaching statistical significance. Further studies with larger sample sizes or biomarker-based stratification may help clarify this potential benefit.

KAPLAN-MEIER SURVIVAL CURVES: Visual representation of patient survival over time, comparing control vs. experimental groups.

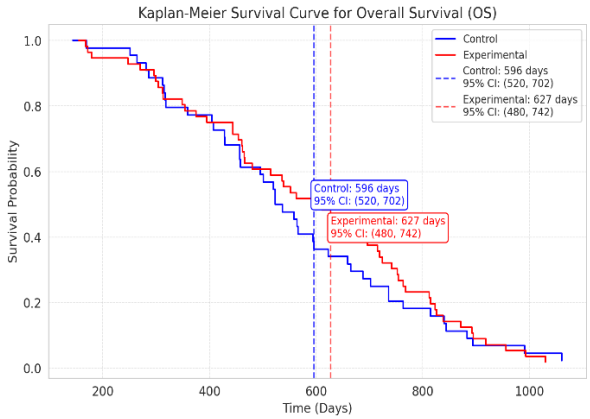

FIG 1: KAPLAN-MEIER SURVIVAL CURVE FOR OS

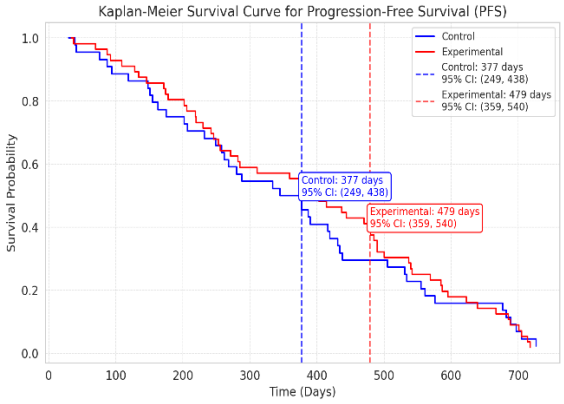

FIG 2: KAPLAN-MEIER SURVIVAL CURVE FOR PFS

INTERPRETATION: The survival curves show that the experimental treatment provides a modest improvement in OS, though statistical significance is not achieved and Similar to OS, PFS trends higher for the experimental group but does not reach statistical significance.

Parameter	Posterior Mean	Posterior SD	95% HPD Interval	Decision
HR (OS)	-0.00271	0.2334	(-0.4645, 0.4015)	Uncertain
HR (PFS)	0.257	0.238	(-0.2041, 0.6961)	Uncertain

TABLE 4: BAYESIAN COX PH MODEL

Table Caption: SD: Standard Deviation and The HPD (Highest Posterior Density) Interval represents the Bayesian credible interval, indicating the range where the true hazard ratio is most likely to fall based on posterior probability distribution, offering a more intuitive uncertainty measure compared to traditional confidence intervals

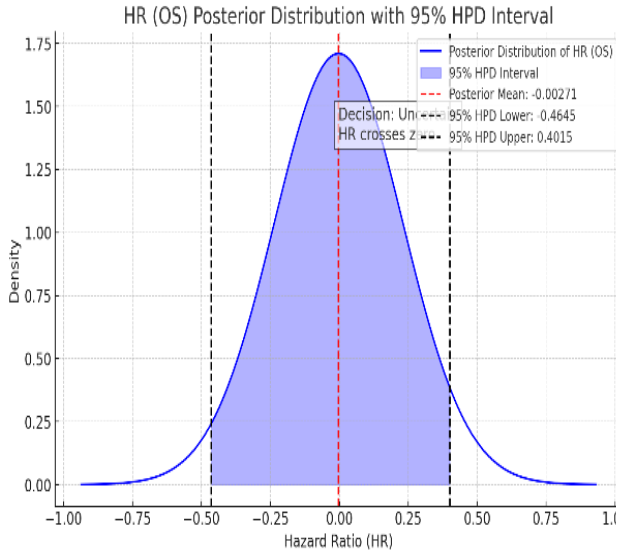

FIG 3: HR POSTERIOR DISTRIBUTIONS OS

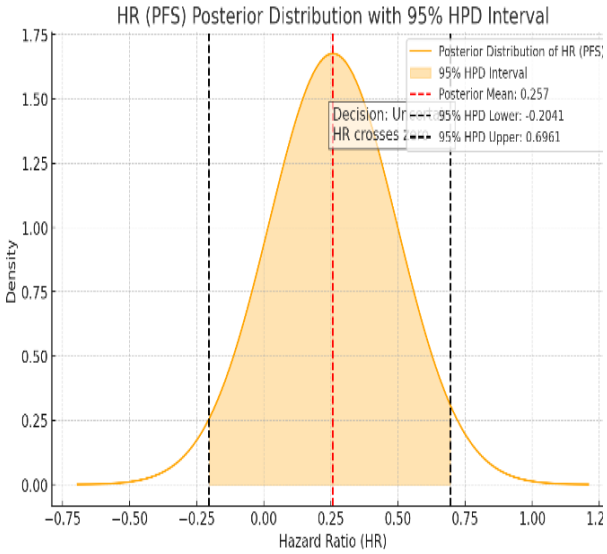

FIG 4: HR POSTERIOR DISTRIBUTIONS PFS

INTERPRETATION: The HR (Hazard Ratio) estimates suggest possible benefits, but high uncertainty in treatment effect (95% HPD includes zero).

Parameter	**Posterior Mean (θ)**	**95% HPD Interval**	**Decision**
OS	0.9984	(0.9954, 1.0000)	Continue
PFS	0.9974	(0.9929, 1.0000)	Continue

TABLE 3: BAYESIAN PREDICTIVE PROBABILITY (Θ):

Table Caption: The HPD (Highest Posterior Density) Interval

INTERPRETATION: The Kaplan-Meier survival analysis did not show a statistically significant difference (log-rank p = 0.9782, Wilcoxon p = 0.9190). However, the Bayesian predictive probability (~99%) suggests that the trial should continue. While Bayesian methods allow real-time probability updates, it is important to note that the SAS PROC MCMC model used in this research does not explicitly account for the treatment variable (TRT01A). Therefore, the predictive probability reflects overall survival trends but does not directly measure the treatment effect. A more reliable assessment of treatment effectiveness would require including the treatment variable in the model.

Bayesian modelling allows continuous learning from accumulating data, helping to evaluate promising treatments over time. However, the claim that it minimizes exposure to ineffective therapies needs further clarification—does the method classify a treatment as ineffective under traditional methods but effective under Bayesian predictive probability? The Bayesian Cox proportional hazards model provides additional insights, with a posterior mean estimate (table 2) of 0.2570 for PFS, suggesting a potential improvement in Progression-Free Survival (PFS) in the experimental group. However, no clear trend was observed for Overall Survival (OS), where the posterior mean estimate was -0. 00271.

BAYESIAN PREDICTIVE PROBABILITY HISTOGRAM: Most Bayesian probabilities indicate high treatment success (~99%), strongly supporting trial continuation.

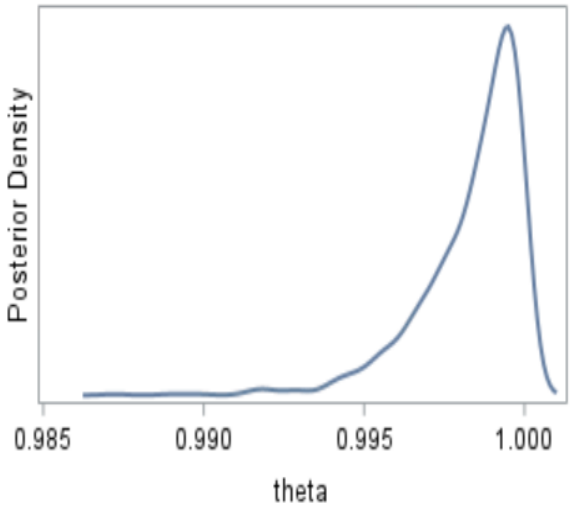

FIG 5: POSTERIOR DENSITY PLOT FOR OS

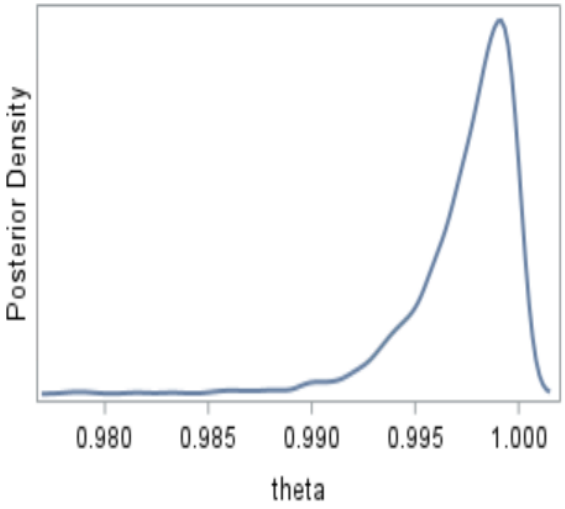

FIG 6: POSTERIOR DENSITY PLOT FOR PFS

INTERPRETATION: The posterior density plots for OS (Figure 5) and PFS (Figure 6) illustrate the Bayesian probability estimates for treatment success. Both distributions are concentrated near 1.0, indicating a high likelihood that the treatment is beneficial. However, it is important to interpret these results cautiously. The Bayesian predictive probability model used in this study does not explicitly account for the treatment variable (TRT01A). As a result, the probability estimates reflect overall survival trends rather than a direct measure of treatment efficacy.

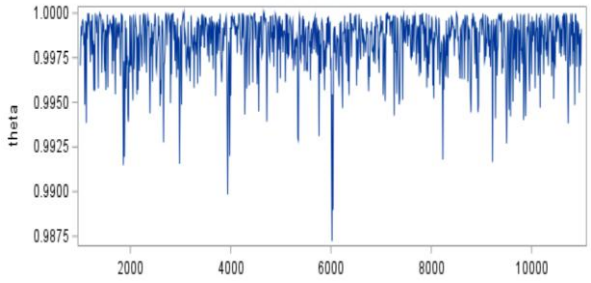

FIG 7: TRACE PLOT (ITERATION VS. THETA) FOR OS

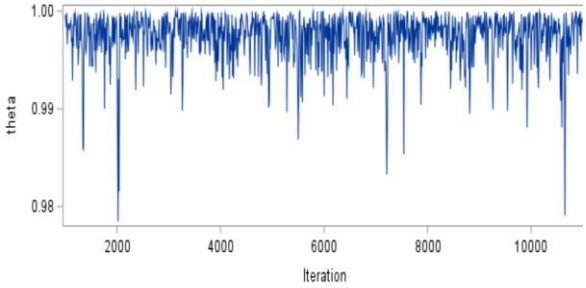

FIG 8: TRACE PLOT (ITERATION VS.THETA) FOR PFS

INTERPRETATION: Since the trace plot remains stable and does not show a systematic trend, we can conclude that the Bayesian inference results are reliable. Most posterior samples align with a high probability of trial continuation.

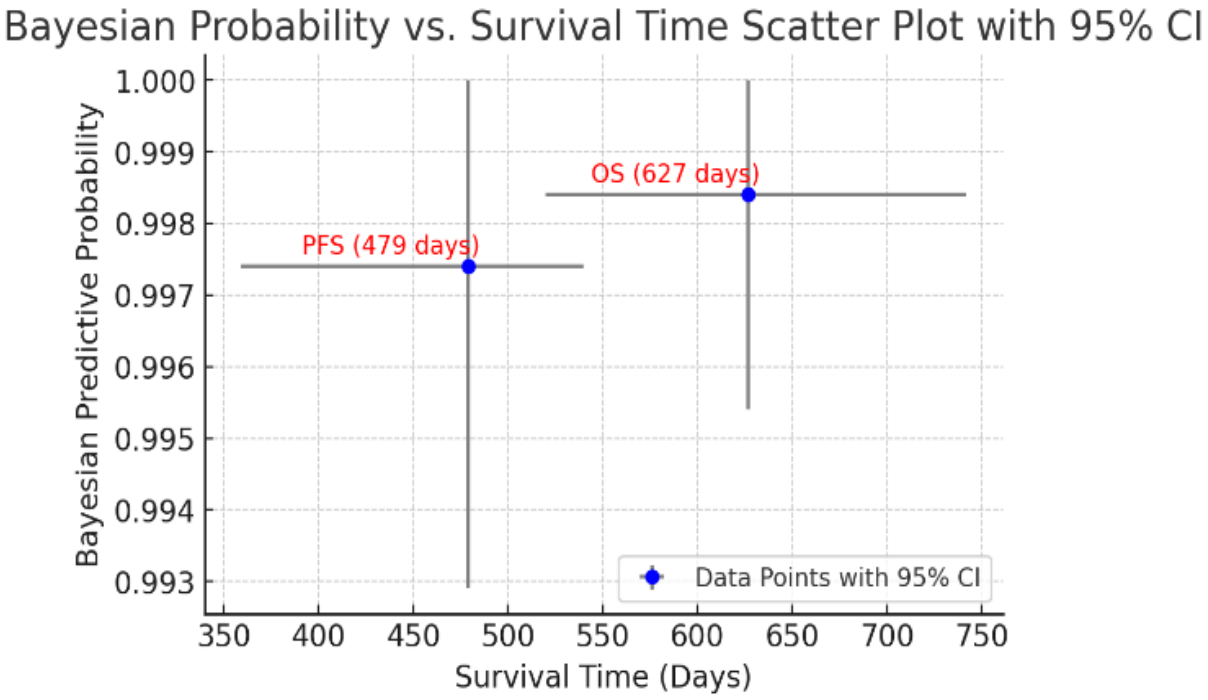

FIG 7: BAYESIAN PROBABILITY VS. SURVIVAL TIME SCATTER PLOT: Bayesian Probability vs. Survival Time Scatter Plot includes 95% Confidence Intervals (CI) for both Bayesian Predictive Probability and Survival Time

Error bars in Gray represent the 95% CI for each value and OS (627 days) and PFS (479 days) as per table 3.

INTERPRETATION: The scatter plot does not show a strong correlation between Bayesian treatment success probability and overall survival (OS) or progression-free survival (PFS). The data points stay widely scattered, with no clear trend or pattern suggesting a meaningful association. Furthermore, while the plot now includes 95% Confidence Intervals (CI) for both Bayesian Predictive Probability and Survival Time, the overlapping ranges further highlight the uncertainty in drawing a definitive correlation.

CONCLUSION

This research demonstrates the advantages of Bayesian adaptive designs in clinical trials for relapsed/refractory Acute Myeloid Leukaemia (AML) by enabling real-time decision-making and trial flexibility. The Bayesian predictive probability (~99%) strongly supports trial continuation, reinforcing its ability to update treatment assessments dynamically.
While Bayesian methods enhance trial efficiency, their role in ensuring patient safety depends on the integration of predefined stopping rules and dose adjustments within the trial design. Additionally, the Bayesian Cox model suggests potential benefits in Progression-Free Survival (PFS), but no significant impact on Overall Survival (OS).

Although the Kaplan-Meier survival analysis did not indicate statistically significant differences, Bayesian inference provides a more dynamic assessment of treatment outcomes. By allowing continuous learning from accumulating data, Bayesian adaptive designs present a promising alternative to traditional survival analysis in improving treatment evaluation.

DISCUSSION

The research demonstrates the potential of Bayesian adaptive designs to optimize cancer trials. While Kaplan-Meier survival curves show modest improvements in OS and PFS for the experimental treatment, the Bayesian Cox PH model indicates uncertainty in treatment effect. However, the Bayesian predictive probability (~99%) strongly supports trial continuation, highlighting the ethical and regulatory advantages of adaptive designs.

REGULATORY CONSIDERATIONS

The research aligns with FDA and EMA guidelines, proving the feasibility of Bayesian adaptive designs in oncology trials. By enabling real-time decision-making, adaptive designs reduce patient exposure to ineffective treatments and accelerate drug development.

LIMITATIONS AND FUTURE WORK

Future work should focus on expanding the sample size and incorporating patient-specific biomarkers (e.g., FLT3, NPM1 mutations) to further refine patient selection and Bayesian adaptive trial designs. Integrating machine learning techniques with Bayesian methods could enhance predictive accuracy and improve the efficiency of adaptive clinical trials. Additionally, further research should investigate hierarchical Bayesian models that consider heterogeneity among patient subgroups, enabling personalized treatment recommendations in adaptive designs.

REFERENCES

- FDA (2019). Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry. U.S. Food and Drug Administration.
- EMA (2007). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design.
- Berry, D. A. (2006). Bayesian clinical trials. Nature Reviews Drug Discovery, 5(1), 27-36.
- Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). Bayesian approaches clinical trials and health-care evaluation. John Wiley & Sons.
- Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). Bayesian survival analysis. Springer.
- Neuenschwander, B., Branson, M., & Gsponer, T. (2008). Critical aspects of Bayesian approaches to clinical trial design. Statistics in Medicine, 27(15), 2420-2439.

ACKNOWLEDGMENTS

I sincerely thank Ephicacy Lifescience Analytics and our team members for their invaluable contributions to this research. Their expertise in Bayesian adaptive trial design, statistical modeling, and clinical data analysis was instrumental in the successful implementation of this research.

Finally, we extend our gratitude to all researchers and collaborators whose constructive feedback greatly contributed to the refinement and advancement of this project.

RECOMMENDED READING

- Chow, S. C., & Liu, J. P. (2008). Design and Analysis of Clinical Trials: Concepts and Methodologies. John Wiley & Sons.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). CRC Press.
- Thall, P. F., Wathen, J. K., Bekele, B. N., Champlin, R. E., Baker, L. H., & Benjamin, R. S. (2003). Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. Statistics in Medicine, 22(5), 763-780.
- O'Quigley, J., Pepe, M., & Fisher, L. (1990). Continual reassessment method: A practical design for phase I clinical trials in cancer. Biometrics, 46(1), 33-48.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: NAVNEET JHA
Company: Ephicacy Lifescience Analytics Pvt, Ltd.,
Address: Bengaluru, India
Work Phone: +91-9510006867
Email: navneet.jha@ephicacy.com
Website: http://www.ephicacy.com

APPENDIX:

Create the AML ADTTE dataset using SAS:

```
/* Step 1: Create the AML ADTTE dataset */
DATA aml_adtte;
    CALL STREAMINIT(42); /* Set seed for reproducibility */
    FORMAT TRTSDT AESTDTC LASTCONTACTDTC DATE9.;

    DO i = 1 TO 100; /* 100 patients, each having 2 records (PFS & OS) */

        /* Unique Subject ID */
        USUBJID = CAT("SUBJ", PUT(i, Z3.));

        /* Assign Treatment Group */
        IF RAND("Uniform") < 0.5 THEN TRT01A = "Experimental";
        ELSE TRT01A = "Control";

        /* Generate Treatment Start Date */
        TRTSDT = '01JAN2020'D + INT(RAND("Uniform") * 365);

        /* Generate PFS Record */
        PARAMCD = "PFS";
        PARAM = "Progression-Free Survival";
        AVAL = ROUND(RAND("Uniform", 90, 730), 1); /* PFS between 3 months - 2 years
*/
        CNSR = (RAND("Uniform") < 0.15); /* 15% censoring */
        EVNTDESC = "Disease Progression";
        AESTDTC = IFN(CNSR = 0, TRTSDT + AVAL, .);
        LASTCONTACTDTC = IFN(CNSR = 1, TRTSDT + INT(RAND("Uniform") * 900), AESTDTC);

        /* Baseline Characteristics */
        Baseline_WBC_Count = ROUND(RAND("Normal", 50, 15), 1);
        Baseline_HGB = ROUND(RAND("Normal", 12, 2), 1);

        /* Assign Cytogenetics Risk */
        Prob_Cyto = RAND("Uniform");
        IF Prob_Cyto < 0.3 THEN Cytogenetics_Risk = "Intermediate";
        ELSE IF Prob_Cyto < 0.8 THEN Cytogenetics_Risk = "Favorable";
        ELSE Cytogenetics_Risk = "Poor";

        /* Assign Mutation Status */
        IF RAND("Uniform") < 0.3 THEN FLT3_Status = "Positive"; ELSE FLT3_Status =
"Negative";
        IF RAND("Uniform") < 0.4 THEN NPM1_Status = "Positive"; ELSE NPM1_Status =
"Negative";

        /* Generate Age */
        Age = ROUND(RAND("Normal", 60, 10), 1);

        /* Prior Treatment */
        Prior_Treatment = (RAND("Uniform") < 0.75); /* 75% prior treatment */

        /* Assign Disease Severity */
        IF RAND("Uniform") < 0.45 THEN Disease_Severity = "Moderate"; ELSE
Disease_Severity = "Severe";

        OUTPUT;

        /* Generate OS Record */
        PARAMCD = "OS";
        PARAM = "Overall Survival";
        AVAL = ROUND(AVAL + RAND("Uniform", 30, 730), 1); /* OS is always greater than
PFS */
        CNSR = (RAND("Uniform") < 0.15);
        EVNTDESC = "Death";
        AESTDTC = IFN(CNSR = 0, TRTSDT + AVAL, .);
```

```
        LASTCONTACTDTC = IFN(CNSR = 1, TRTSDT + INT(RAND("Uniform") * 900), AESTDTC);

        OUTPUT;
    END;

    DROP i Prob_Cyto;
RUN;


/* Step 3: Export Dataset to CSV */
PROC EXPORT DATA=aml_adtte
    OUTFILE="/Desktop/Phuse/AML_ADTTE_rand.csv"
    DBMS=CSV REPLACE;
RUN;

   /* Step 1: Ensure Proper Labeling */
DATA aml_adtte;
retain USUBJID TRT01A TRTSDT PARAMCD PARAM AVAL CNSR EVNTDESC AESTDTC LASTCONTACTDTC
Baseline_WBC_Count Baseline_HGB Cytogenetics_Risk FLT3_Status NPM1_Status;

    SET aml_adtte;
    LABEL
        USUBJID = "Unique Subject Identifier"
        TRT01A = "Treatment Group (Control / Experimental)"
        TRTSDT = "Treatment Start Date"
        PARAMCD = "Parameter Code (PFS, OS)"
        PARAM = "Parameter Name"
        AVAL = "Time to Event (Days)"
        CNSR = "Censoring Indicator (0 = Event, 1 = Censored)"
        EVNTDESC = "Event Description (Progression, Death, Censored)"
        AESTDTC = "Date of First Event (Disease Progression or Death)"
        LASTCONTACTDTC = "Last Known Contact Date"
        Baseline_WBC_Count = "White Blood Cell Count (10^9/L) at Treatment Start"
        Baseline_HGB = "Hemoglobin Level (g/dL) at Treatment Start"
        Cytogenetics_Risk = "AML Risk Classification: Favorable, Intermediate, Poor"
        FLT3_Status = "FLT3 Mutation Status: Positive / Negative"
        NPM1_Status = "NPM1 Mutation Status: Positive / Negative";

RUN;
```