PharmaSUG 2025 - Paper ST-231

Mastering the Maze of Oncology Endpoints: A Unified SAS Approach for Randomized Controlled Trial Analysis

Yuxin Wang, Kelly Chao, Wenqing Yu and Hongbing Jin. LLX Solutions, LLC

ABSTRACT

In oncology clinical trials, particularly in randomized controlled studies, the analysis of diverse endpoints is crucial for evaluating treatment efficacy compared to the control arm. This paper presents a comprehensive, unified approach to analyzing three fundamental types of endpoints in oncology research: time-to-event, categorical and continuous. By consolidating these methodologies into a single, accessible resource, we aim to provide an invaluable reference for statisticians, programmers, and researchers in the field.

Our work focuses on three key areas of statistical analysis, each tailored to a specific endpoint type.

Time-to-Event Endpoints: We examine both non-parametric (Kaplan-Meier) and semi-parametric (Cox proportional hazards regression) methods for analyzing endpoints like overall survival (OS) and progression-free survival (PFS). The paper introduces PROC LIFETEST for survival curve generation and estimation and PROC PHREG for Cox regression and model evaluation.

Categorical Endpoints: We explore methods for evaluating endpoints such as Objective Response Rate (ORR) and Disease Control Rate (DCR). We detail statistical tests for comparing response rates including Cochran-Mantel-Haenszel (CMH) test, Chi-square test, and Fisher's exact test. This includes calculating confidence intervals for response rates and between-group differences using PROC FREQ with the RISKDIFF option.

Continuous Endpoints: We elaborate on the Mixed Model Repeated Measures (MMRM) approach for analyzing continuous endpoints like functional tests in patient-reported outcomes (PRO). This section addresses the critical issue of missing data in longitudinal studies, demonstrating the use of PROC MIXED and PROC MI for data imputation under missing at random (MAR) and missing not at random (MNAR) assumptions.

INTRODUCTION

In oncology clinical trials, the selection and analysis of appropriate endpoints are critical for evaluating the efficacy of new treatments. These endpoints serve as objective measures of how a medical intervention benefits a patient's feeling, function, and survival. As the landscape of cancer research evolves, the importance of diverse endpoints in assessing treatment outcomes has become increasingly apparent.

Oncology trials typically employ three fundamental types of endpoints: categorical, continuous, and time-to-event. Categorical endpoints, such as Objective Response Rate (ORR) and Disease Control Rate (DCR), provide discrete measures of treatment response. Continuous endpoints, like functional tests in patient-reported outcomes (PROs), offer nuanced insights into treatment effects over time. Time-to-event endpoints, including the gold standard Overall Survival (OS) and the surrogate Progression-Free Survival (PFS), assess the duration of treatment benefit.

While Overall Survival remains the most meaningful clinical endpoint in cancer research and clinical trials, the use of surrogate endpoints has gained traction due to their ability to expedite the evaluation of potentially life-saving treatments. This shift reflects the need to balance rigorous scientific assessment with the urgency of bringing innovative therapies to patients.

Despite the wealth of literature on individual endpoint types, there is a notable lack of comprehensive resources that unify the statistical approaches for analyzing these diverse endpoints in comparing the treatment and control group. This paper aims to address this gap by providing a consolidated, accessible reference for statisticians, programmers, and researchers in the field of oncology. By offering a unified approach to endpoint analysis, we seek to enhance the efficiency and robustness of clinical trial data

analysis and interpretation, ultimately contributing to more informed decision-making in cancer treatment development.

TYPICAL WAYS TO EVALUATE THE TIME-TO-EVENT ENDPOINTS

KAPLAN-MEIER

The Kaplan-Meier method is a cornerstone in analyzing various time-to-event endpoints in oncology studies. It is particularly useful for evaluating overall survival (OS), which measures the time from randomization or treatment initiation to death from any cause. Progression-free survival (PFS), another critical endpoint, uses K-M analysis to assess the time until disease progression or death. Event-free survival (EFS) extends this concept to include other clinically significant events. Duration of response (DOR) applies K-M methodology to responders, measuring the time from initial response to disease progression or death. These endpoints provide comprehensive insights into treatment efficacy and patient outcomes.

One of the strengths of the K-M method is its ability to handle censored data effectively. In oncology trials, censoring occurs when the event of interest (e.g., death or disease progression) is not observed for some patients during the study period. This can happen due to patient drop-out, loss to follow-up, or study conclusion before all events occur. K-M analysis incorporates these censored observations, providing unbiased estimates of survival probabilities and maximizing the use of available data, which is crucial in studies where patient retention can be challenging.

K-M analysis is crucial for estimating median survival times, a key statistic in oncology trials. Median survival is the time point at which the survival probability drops to 50%. This metric provides a tangible, easily interpretable measure of treatment effect. In cases where the survival curve does not drop below 50% during the study period, K-M analysis allows for the estimation of landmark survival rates at specific time points (e.g., 1-year or 5-year survival rates), offering valuable insights into long-term treatment outcomes.

Note: The minimum and maximum survival time can be considered as the cases when survival probability close to 1 or 0. Minimum survival time is the shortest time that the event of interest is observed. Maximum survival time is the longest observed time among all subjects.

K-M plots are powerful visual tools in oncology research. These step-function graphs display the estimated probability of survival over time for different treatment groups. The x-axis typically represents time since treatment initiation or randomization, while the y-axis shows the proportion of patients surviving. These plots allow for immediate visual comparison of survival patterns between treatment arms, highlighting differences in long-term outcomes. Researchers can easily identify if one treatment consistently outperforms another or if survival curves cross, indicating time-dependent treatment effects.

We can also use KM method to determine the statistical significance of survival differences between treatment groups. The log-rank test, frequently used in conjunction with K-M analysis, is a powerful tool for determining the statistical significance of survival differences between treatment groups. This non-parametric test compares the entire survival experience of two or more groups, considering the timing of events throughout the follow-up period. It provides a p-value indicating whether observed differences in survival curves are likely to be due to chance or represent true treatment effects. The log-rank test complements the visual assessment of K-M curves, offering a quantitative measure of treatment efficacy that is crucial for decision-making in clinical oncology.

The code below is typically used in oncology studies to analyze time-to-event data, compare survival between treatment groups, and estimate survival probabilities at specific time points.

```
ods output censoredsummary = censor productlimitestimates = est quartiles = quart
homtests = hom;
proc lifetest data = cs_tte alpha=0.05 timelist = 70 80 90 100 110 120 outsurv=sur;
strata group_no;
time aval*cnsr(1);
run;
```

Program 1. Use PROC LIFETEST to use KM method to analyze the time-to-event endpoint

The SAS code serves several purposes in survival analysis using PROC LIFETEST:

- ODS Output: The 'ods output' statement creates three output datasets:

  'censor': Contains summary statistics for censored observations

  'quart': Stores quartile estimates of survival time

  'est': Holds the product-limit (Kaplan-Meier) estimates of the survivor function

  'hom': Hypothesis test results of Log-Rank test.

  Options:

  • 'timelist = 70 80 90 100 110 120': Specifies time points (in months) for survival probability estimation

  • 'alpha=0.05': Sets the significance level for computing confidence intervals of the quartiles to 5%

  • 'outsurv=sur': Creates an output dataset 'sur' containing the survival estimates

- STRATA statement: Divides the data into strata based on the 'group_no', allowing for comparison of survival curves between groups
- TIME statement: Specifies the time variable 'aval' and the censoring variable 'cnsr', where 1 indicates censored observations

Below is the screenshot of each result from the ODS OUTPUTS statement.

VIEWTABLE: Work.Censor (Censored Summary)

	STRATUM	GROUP_NO	TOTAL	FAILED	CENSORED	PCTCENS
1	1	2	16	15	1	6.25
2	2	3	22	22	0	0.00
3	3	4	16	15	1	6.25
4	4	5	16	15	1	6.25
5	5	99	54	52	2	3.70

Figure 1. SAS Dataset: summary statistics for censored observations

VIEWTABLE: Work.Quart (Quartiles of the Survival Distribution)

	STRATUM	GROUP_NO	PERCENT	ESTIMATE	TRANSFORM	LOWERLIMIT	UPPERLIMIT
1	1	2	75	12.0000	LOGLOG	8.0000	
2	1	2	50	8.0000	LOGLOG	6.0000	11.0000
3	1	2	25	7.0000	LOGLOG	6.0000	7.0000
4	2	3	75	11.0000	LOGLOG	8.0000	19.0000
5	2	3	50	8.0000	LOGLOG	5.0000	10.0000
6	2	3	25	5.0000	LOGLOG	4.0000	7.0000
7	3	4	75	21.0000	LOGLOG	15.0000	
8	3	4	50	15.0000	LOGLOG	11.0000	21.0000
9	3	4	25	11.0000	LOGLOG	7.0000	15.0000
10	4	5	75	14.0000	LOGLOG	10.0000	
11	4	5	50	11.0000	LOGLOG	8.0000	14.0000
12	4	5	25	8.5000	LOGLOG	4.0000	10.0000
13	5	99	75	15.0000	LOGLOG	13.0000	19.0000
14	5	99	50	10.0000	LOGLOG	8.0000	13.0000
15	5	99	25	7.0000	LOGLOG	5.0000	8.0000

Figure 2. SAS Dataset of the quartile estimates of survival time

VIEWTABLE: Work.Est (Product-Limit Estimates)

	STRATUM	PARAMCD	GROUP_NO	TIMELIST	AVAL	CENSOR	SURVIVAL	FAILURE	STDERR	FAILED	LEFT
1	1	DURSC2	2	3.0000	0.0000	0	1.0000	0	0	0	16
2	1	DURSC2	2	5.0000	0.0000	0	1.0000	0	0	0	15
3	1	DURSC2	2	7.0000	7.0000	0	0.5333	0.4667	0.1288	7	8
4	1	DURSC2	2	9.0000	9.0000	0	0.4000	0.6000	0.1265	9	6
5	1	DURSC2	2	11.0000	11.0000	0	0.2667	0.7333	0.1142	11	4
6	1	DURSC2	2	13.0000	12.0000	0	0.1333	0.8667	0.0878	13	2
7	1	DURSC2	2	15.0000	14.0000	0	0	1.0000		15	0
8	1	DURSC2	2	17.0000	14.0000	0	0	1.0000		15	0
9	1	DURSC2	2	19.0000	14.0000	0	0	1.0000		15	0
10	1	DURSC2	2	21.0000	14.0000	0	0	1.0000		15	0
11	1	DURSC2	2	23.0000	14.0000	0	0	1.0000		15	0
12	2	DURSC2	3	3.0000	0.0000	0	1.0000	0	0	0	22
13	2	DURSC2	3	5.0000	5.0000	0	0.7273	0.2727	0.0950	6	16
14	2	DURSC2	3	7.0000	7.0000	0	0.5455	0.4545	0.1062	10	12
15	2	DURSC2	3	9.0000	8.0000	0	0.3182	0.6818	0.0993	15	7
16	2	DURSC2	3	11.0000	11.0000	0	0.2273	0.7727	0.0893	17	5
17	2	DURSC2	3	13.0000	12.0000	0	0.1364	0.8636	0.0732	19	3
18	2	DURSC2	3	15.0000	15.0000	0	0.0909	0.9091	0.0613	20	2
19	2	DURSC2	3	17.0000	15.0000	0	0.0909	0.9091	0.0613	20	2
20	2	DURSC2	3	19.0000	19.0000	0	0.0455	0.9545	0.0444	21	1
21	2	DURSC2	3	21.0000	19.0000	0	0.0455	0.9545	0.0444	21	1
22	2	DURSC2	3	23.0000	19.0000	0	0.0455	0.9545	0.0444	21	1

Figure3. SAS Dataset: product-limit (Kaplan-Meier) estimates of the survivor function

VIEWTABLE: Work.Hom (Homogeneity Tests)

	TEST	CHISQ	DF	PROBCHISQ
1	Log-Rank	154.6828	9	<.0001
2	Wilcoxon	145.5100	9	<.0001
3	-2Log(LR)	43.8674	9	<.0001

Figure4. SAS Dataset of the log-rank test result

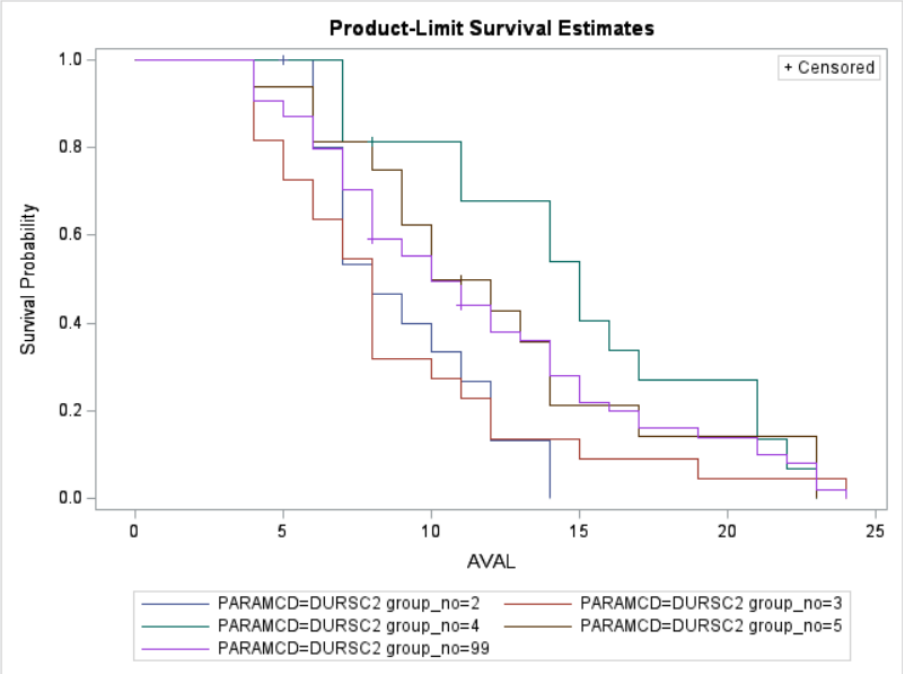


Figure5. KM Plot

COX PROPORTIONAL HAZARDS

The Cox Proportional Hazards model is a widely used statistical method in survival analysis for evaluating the effect of multiple risk factors on survival time. It is a semi-parametric model that estimates hazard ratios without requiring specification of the baseline hazard function, making it flexible and robust. The model's key assumption is proportional hazards, which implies that the hazard ratio between groups remains constant over time. Its hazard function is expressed as below.

$$h(t) = h_0(t) \times \exp(b_1x_1 + b_2x_2 + \cdots + b_px_p)$$

The primary output of the model, hazard ratios (HR), provides an intuitive interpretation by representing the relative risk of an event occurring associated with each covariate. This makes the Cox model a powerful tool for analyzing time-to-event data while accounting for multiple predictors simultaneously.

Here's an example of how to implement the Cox Proportional Hazards model using SAS PROC PHREG:

```
PROC PHREG DATA=your_dataset;
  CLASS categorical_var1 categorical_var2;
  MODEL time*status(0) = continuous_var1 categorical_var1 categorical_var2;
  WHERE condition;
RUN;
```

Program 2. Use PROC PHREG to use Cox Proportional Hazards model to analyze the time-to-event endpoint

Key components of the SAS code:

- PROC PHREG: Calls the procedure for Cox regression.
- DATA: Specifies the input dataset.
- CLASS: Identifies categorical variables.
- MODEL: Defines the time variable, censoring indicator (status), and covariates.
- WHERE: Optional statement to filter data.

This code will produce hazard ratios, p-values, and confidence intervals for each covariate, allowing researchers to assess their impact on survival

COMPARISON BETWEEN THE KAPLAN-MEIER AND COX PROPORTIONAL HAZARDS MODEL

The Kaplan-Meier method and the Cox Proportional Hazards model are two widely used approaches in survival analysis, each serving distinct purposes. The Kaplan-Meier method focuses on estimating and visualizing survival probabilities over time without adjusting for covariates, making it ideal for descriptive analyses. In contrast, the Cox Proportional Hazards model is designed to analyze the effect of multiple risk factors on survival simultaneously, offering a more comprehensive understanding of how various predictors influence survival outcomes.

A key difference between the two methods lies in their handling of covariates. The Kaplan-Meier method is limited in that it cannot incorporate multiple variables or continuous risk factors, restricting its application to group-level comparisons. On the other hand, the Cox Proportional Hazards model adjusts for multiple risk factors and accommodates continuous variables, providing a more nuanced analysis of survival data.

The output of these methods also differs significantly. The Kaplan-Meier method produces survival curves and estimates median survival times, which are useful for visualizing and summarizing survival trends. In contrast, the Cox Proportional Hazards model provides hazard ratios, confidence intervals, and p-values for each covariate, enabling researchers to quantify the relative risk associated with specific predictors and assess their statistical significance.

Both methods rely on specific assumptions. The Kaplan-Meier method assumes independent observations and non-informative censoring, ensuring that censored data do not bias the results. The Cox

Proportional Hazards model builds on these assumptions but also requires the proportional hazards assumption, which states that hazard ratios remain constant over time.

In terms of flexibility, the Kaplan-Meier method is limited to comparing survival between groups, making it less suitable for complex analyses. In contrast, the Cox Proportional Hazards model offers greater flexibility by allowing researchers to analyze complex relationships and interactions between variables, making it a powerful tool for investigating multifactorial influences on survival outcomes.
While both methods can handle right-censored data, the Cox model is more versatile for multivariable analysis in survival studies. Researchers often use Kaplan-Meier for initial survival curve visualization and then proceed to Cox regression for more comprehensive analysis of risk factors.

CATEGORICAL ENDPOINTS ANALYSIS

In oncology clinical trials, categorical endpoints such as Objective Response Rate (ORR) serve as critical measures of treatment efficacy, often reflecting the proportion of patients achieving predefined therapeutic responses. Comparing response rates between treatment arms requires robust statistical methods to account for variability, sample size, and potential confounding factors. Confidence intervals for the difference in response rates provide a measure of precision, complementing hypothesis testing by quantifying uncertainty around effect estimates. Three widely used methods for such comparisons include the Cochran-Mantel-Haenszel (CMH) test, Chi-square test, and Fisher's exact test. When choosing between these three statistical tests for categorical data analysis, it's important to understand their distinct characteristics and use cases.

The Cochran-Mantel-Haenszel (CMH) test is specifically designed for comparing two binary variables while controlling for a third variable through stratification. It accounts for confounding factors by analyzing the association across multiple strata and is more powerful than separate chi-square tests for each stratum, making it particularly useful in clinical trials when adjusting for site effects or other stratification factors.

The Chi-Square test, on the other hand, is best suited for large sample sizes (typically n > 30) and tests independence between two categorical variables. It requires that expected frequencies should be greater than 5 in each cell and that observations are independent, though it becomes less accurate with small sample sizes.

The Fisher's Exact test is preferred for small sample sizes, especially when cell counts are less than 5, as it calculates the exact probability of observing the results. While more computationally intensive than the chi-square test, it's more conservative and less likely to reject the null hypothesis, making it particularly valuable for analyzing rare adverse events in small patient populations.

Overall, when selecting between these tests, use the CMH test if you have stratification factors, Fisher's exact test for small samples or rare events, and the chi-square test for large sample sizes with adequate cell counts.

The PROC FREQ is the main procedure used for all three tests, with different options for each test type.

We provide the sample code below to use the PROC FREQ to calculate the difference in Objective Response Rate (Treatment vs Placebo) with 95% CI, and the corresponding P-value.

```
ods output CMH =pvalue CommonPdiff=diff(where=(method in ('Mantel-Haenszel')));
proc freq data=trial_data;
      tables center*treatment*response/cmh COMMONRISKDIFF;
run;
```

Program 3. Use PROC FREQ to calculate CMH test P-value and the difference in ORR with 95% in Mantel-Haenszel CI

The results are listed below.

	TABLE	STATISTIC	ALTHYPOTHESIS	DF	VALUE	PROB
1	Summary for TRTN * RES	1	Nonzero Correlation	1	20.5837	<.0001
2	Summary for TRTN * RES	2	Row Mean Scores Differ	1	20.5837	<.0001
3	Summary for TRTN * RES	3	General Association	1	20.5837	<.0001

Figure6. SAS Dataset: CMH test P-value

	TABLE	COLUMN	METHOD	VALUE	STDERR	LOWERCL	UPPERCL
1	Summary for TRTN * RES	1	Mantel-Haenszel	0.3914	0.0544	0.2848	0.4979

Figure7. SAS Dataset: the difference in ORR with 95% in Mantel-Haenszel CI

```
proc freq data = trial_data;
tables treatment*response/riskdiff(cl=(wald)) chisq ;
ods output RiskDiffCol1 = b0 chisq=chi(where=(STATISTIC='Chi-Square')) CrossTabFreqs=f1;
run;
```

Program 4. Use PROC FREQ to calculate Chi-square test P-value and the difference in ORR with 95% Unstratified Wald CI

The results are listed below.

VIEWTABLE: Work.Chi (Chi-Square Tests)

	TABLE	STATISTIC	DF	VALUE	PROB
1	Table TRTN * RES	Chi-Square	1	21.7349	<.0001

Figure6. SAS Dataset: Chi-Square test P-value

VIEWTABLE: Work.B0 (Column 1 Risk Estimates)

	TABLE	ROW	RISK	ASE	LOWERCL	UPPERCL	EXACTLOWERCL	EXACTUPPERCL	CONTROL
1	Table TRTN * RES	Row 1	0.3976	0.0537	0.2923	0.5029	0.2917	0.5110	
2	Table TRTN * RES	Row 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0881	
3	Table TRTN * RES	Total	0.2683	0.0400	0.1900	0.3466	0.1924	0.3557	
4	Table TRTN * RES	Difference	0.3976	0.0537	0.2923	0.5029	_	_	1

Figure7. SAS Dataset: the difference in ORR with 95% Unstratified Wald CI

```
proc freq data = trial_data;
   tables treatment*response/riskdiff(cl=(exact)) fisher ;
   exact riskdiff;
   ods output pdiffcls = b2 fishersexact=fish(where=(label1='Two-sided Pr <= P'))
CrossTabFreqs=f2;
run;
```

Program 5. Use PROC FREQ to calculate Fisher Exact test P-value and the difference in ORR with 95% Unstratified Exact CI

The results are listed below.

VIEWTABLE: Work.Fish (Fisher's Exact Test)

	TABLE	NAME1	LABEL1	CVALUE1	NVALUE1
1	Table TRTN * RES	XP2_FISH	Two-sided Pr <= P	<.0001	0.000000182

Figure6. SAS Dataset: Fisher Exact test P-value

VIEWTABLE: Work.B2 (Risk Difference Confidence Limits)

	TABLE	COLUMN	RISKDIFFERENCE	TYPE	LOWERCL	UPPERCL
1	Table TRTN * RES	1	0.3975903614	Exact (Score)	0.2882	0.5110

Figure7. SAS Dataset: the difference in ORR with 95% Unstratified Exact CI

MMRM MODEL TO EVALUATE THE CONTINUOUS ENDPOINTS

In oncology clinical trials, continuous endpoints derived from patient-reported outcomes (PROs) have become increasingly vital for capturing treatment effects on symptoms, functional status, and health-related quality of life (HRQoL). PRO tools such as the EQ-5D, FACT-G, and disease-specific instruments like the EORTC QLQ-C30 or NSCLC-SAQ provide longitudinal, continuous measurements of patient experiences, offering insights beyond traditional survival metrics. These endpoints are particularly valuable for evaluating interventions in cancers with prolonged survival timelines, where maintaining or improving daily function and symptom control are critical therapeutic goals. However, analyzing such data presents challenges due to inherent complexities like repeated measurements over time, within-subject correlations, and missing data patterns arising from patient dropout or irregular assessments.

The Mixed Models for Repeated Measures (MMRM) framework addresses these challenges through its robust statistical design. MMRM is specifically tailored for longitudinal continuous data, accommodating repeated measurements on the same subject while accounting for within-subject correlation structures, such as autoregressive or unstructured covariance patterns. This approach models both the mean trajectory of outcomes and the covariance structure simultaneously, enhancing precision in treatment effect estimation. A key strength of MMRM lies in its ability to handle missing data under the Missing at Random (MAR) assumption, reducing bias compared to traditional methods like last-observation-carried-forward.

PROC MIXED in SAS can calculate the LS Mean and the difference in LS Mean of change from baseline between treatment and placebo with the corresponding 95% confidence interval and the p-value for a specific timepoint (avisitn in the code below).

MMRM Model (Treatment vs. Placebo)	**Treatment**	**Placebo**
LS Mean (SE)	XX (XX)	XX (XX)
95% CI of LS Mean	(XX,XX)	(XX,XX)
Difference (95% CI) in LS Mean	XX (XX, XX)	
p-value	XX	

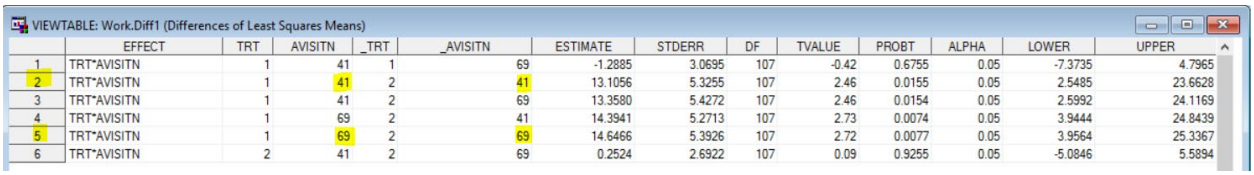
Table 1. Data display format of application of MMRM Model

Below is the SAS code to get the results from the table.

```
proc mixed data=trial_data(where=(chg^=.)) method=reml covtest empirical
plots(maxpoints=none);
      class treatment avisitn group usubjid;
      model chg = treatment avisitn treatment*avisitn group base/ddfm=bw cl;
      repeated avisitn / subject=usubjid type=un r;
      lsmeans treatment*avisitn/diff=all cl e alpha=.05;
      ods output diffs= diff1 LSMeans= LSMean1;
run;
```

Program 6. Use PROC MIXED to use MMRM model to analyze the continuous endpoint

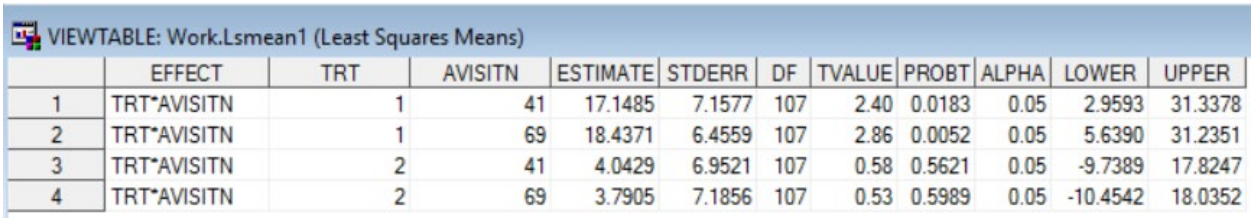
The results are listed below.



VIEWTABLE: Work.Diff1 (Differences of Least Squares Means)

	EFFECT	TRT	AVISITN	_TRT	_AVISITN	ESTIMATE	STDERR	DF	TVALUE	PROBT	ALPHA	LOWER	UPPER
1	TRT*AVISITN	1	41	1	69	-1.2885	3.0695	107	-0.42	0.6755	0.05	-7.3735	4.7965
2	TRT*AVISITN	1	41	2	41	13.1056	5.3255	107	2.46	0.0155	0.05	2.5485	23.6628
3	TRT*AVISITN	1	41	2	69	13.3580	5.4272	107	2.46	0.0154	0.05	2.5992	24.1169
4	TRT*AVISITN	1	69	2	41	14.3941	5.2713	107	2.73	0.0074	0.05	3.9444	24.8439
5	TRT*AVISITN	1	69	2	69	14.6466	5.3926	107	2.72	0.0077	0.05	3.9564	25.3367
6	TRT*AVISITN	2	41	2	69	0.2524	2.6922	107	0.09	0.9255	0.05	-5.0846	5.5894

Figure8. SAS Dataset: the difference of Least Squares Means in AVISITN 41 and 69 between the treatment and placebo groups and the p-value (PROBT)



VIEWTABLE: Work.Lsmean1 (Least Squares Means)

	EFFECT	TRT	AVISITN	ESTIMATE	STDERR	DF	TVALUE	PROBT	ALPHA	LOWER	UPPER
1	TRT*AVISITN	1	41	17.1485	7.1577	107	2.40	0.0183	0.05	2.9593	31.3378
2	TRT*AVISITN	1	69	18.4371	6.4559	107	2.86	0.0052	0.05	5.6390	31.2351
3	TRT*AVISITN	2	41	4.0429	6.9521	107	0.58	0.5621	0.05	-9.7389	17.8247
4	TRT*AVISITN	2	69	3.7905	7.1856	107	0.53	0.5989	0.05	-10.4542	18.0352

Figure9. SAS Dataset: the Least Squares Means in AVISITN 41 and 69 of treatment and placebo groups

The choice of covariance structure in MMRM models is crucial for obtaining valid results. The Unstructured (UN) covariance is the most flexible as it estimates unique variance and covariance for each time point without pattern assumptions. While it provides the best fit for large samples, it requires more parameters and may have convergence issues with smaller samples. Compound Symmetry (CS) assumes constant variance and equal correlations between time points, making it simpler and suitable for small samples, but it may be too restrictive for real data. The Autoregressive First Order (AR(1)) structure assumes correlations decrease exponentially with time, making it appropriate for equally spaced measurements in longitudinal studies, though it may not fit well with unequal visit spacing. The Toeplitz (TOEP) structure allows different correlations for each time lag while maintaining the same correlation for equal time differences, offering more flexibility than AR(1) but requiring more parameters.

When selecting a covariance structure, it's recommended to start with UN if the sample size is sufficient, then compare different structures using fit statistics like AIC/BIC. The final choice should consider factors such as sample size, number of time points, visit spacing, and computational resources. These structures can be compared in SAS using PROC MIXED with different TYPE= specifications in the REPEATED statement, along with information criteria to guide the selection of the most appropriate structure for your specific data.

If we would like to apply the MMRM model under the MNAR (missing not at random) assumption, the PROC MI can be used to do multiple imputation method with the control based. The steps are as below.

Step 1: Before the imputation, the dataset needs to be organized into a horizontal structure with the columns of subject id, treatment, strata, and the observed values of each timepoint.

Step 2: Use the code below to construct the dataset with each subject repeated 100 times.

```
proc mi data=trial out=trial_mono seed=20250329 nimpute=100;
   mcmc chain=multiple impute=monotone displayinit initial=em(itprint);
   var trt strata1 strata2 base visit1 visit2;
run;
```

Program 7. Use PROC MI to repeat 100 times

Step 3: Use the code below to impute the missing data based on the trt 2 with the MNAR statement.

```
proc mi data=trial_mono out= trial_mono_2 nimpute=1 seed=20250329;
   by _imputation_;
   class trt;
   monotone reg(visit1 visit2/details);
   var strata1 strata2 base visit1 visit2;
   mnar model(visit1 visit2/modelobs=(trt='2'));
run;
```

Program 8. Use PROC MI to impute the missing data with control based

Step 4: Derive the change from baseline based on the imputed data.

Step 5: Apply the MMRM model using PROC MIXED based on the previous sections discussed. One thing to notice is that the by statement is included because of the multiple imputation.

```
ods output diffs= diffs1(where=(_trt=2 and avisitn=_avisitn)) LSMeans= lsm1;
proc mixed data=modelchg(where=(chg^=.)) method=reml covtest empirical
plots(maxpoints=none);
    by _imputation_;
    class trt(ref='2') avisitn strata1 strata2 usubjid;
    model chg = trt avisitn trt*avisitn strata1 strata2 base/ddfm=bw cl;
    repeated avisitn / subject=usubjid type=&type r;
    lsmeans trt*avisitn/pdiff cl e alpha=.05;
run;
```

Program 9. Use PROC MIXED to analyze the multiple imputation based on MMRM model

Step 6: Use the PROC MIANALYZE to combine the MMRM results of multiple imputation

```
proc mianalyze parms(classvar=full)=lsm1;
   class trt avisitn;
   modeleffects trt*avisitn;
   ods output parameterestimates=lsm1_;
run;

proc mianalyze parms(classvar=full)=diffs1;
   class trt avisitn;
   modeleffects trt*avisitn;
   ods output parameterestimates=diffs1_;
run;
```

Program 10. Use PROC MIANALYZE to combined multiple imputation MMRM analysis results

The tipping point sensitivity analysis is also commonly used to find the recovery point from significant to not significant based on the multiple imputations. Repeat the code below for each shift value. And then repeat the step 4 to 6 mentioned above.

```
proc mi data=trial_mono seed=20250329 out=data_tip1 nimpute=1;
      by _Imputation_;
      class trt;
      monotone reg(visit1 visit2/details);
      var trt strata1 strata2 visit1 visit2;
      mnar adjust(visit1 / shift=value1 adjustobs=(trtpn ='1') );
      mnar adjust(visit2 / shift=value1 adjustobs=(trtpn ='1') );
run;
```

Program 11. Use PROC MI to do sensitivity tipping point analysis

CONCLUSION

This paper provides a unified framework for analyzing the three fundamental types of endpoints in oncology clinical trials—time-to-event, categorical, and continuous—offering a comprehensive resource for statisticians, programmers, and researchers

By consolidating these methodologies into a single reference, this paper bridges a critical gap in oncology research by equipping researchers with practical tools for analyzing diverse endpoints. The integration of these statistical approaches enhances the rigor and interpretability of clinical trial data, facilitating more informed decision-making in cancer treatment development.

Ultimately, this unified approach underscores the importance of tailoring statistical methods to endpoint-specific characteristics while maintaining flexibility to address real-world challenges in oncology trials. This work aims to serve as a foundational guide for advancing robust and efficient analyses in statistical cancer analysis.

REFERENCES

Savithri Jajam, Vamshi Matta, Lavanya Peddibhotla. 2021. “Confused with PROC MIXED? Find Out What is the Best Model.” Accessed January 7, 2025.
https://www.lexjansen.com/phuse-us/2021/as/PAP_AS09.pdf
Chris Smith, Scott Kosten. 2017. “Multiple Imputation: A Statistical Programming Story.” Accessed January 10, 2025.
https://pharmasug.org/proceedings/2017/SP/PharmaSUG-2017-SP01.pdf
National Library of Medicine. 2023. “A multistate platform model for time to event endpoints in oncology clinical trials.” Accessed January 10, 2025.
https://pmc.ncbi.nlm.nih.gov/articles/PMC10787202/
Joy Zeng. 2019. “Useful SAS techniques in Efficacy Analysis for Oncology studies.” Accessed January 10, 2025.
https://pharmasug.org/proceedings/2019/BP/PharmaSUG-2019-BP-084.pdf

CONTACT INFORMATION

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

Name: Yuxin Wang
Enterprise: LLX Solutions, LLC
Email: yuxin.wang@llxsolutions.com

Name: Kelly Chao
Enterprise: LLX Solutions, LLC
Email: kelly.chao@llxsolutions.com

Name: Wenqing Yu
Enterprise: LLX Solutions, LLC
Email: wenqing.yu@llxsolutions.com

Name: Hongbing Jin
Enterprise: LLX Solutions, LLC
Email: hongbing.jin@llxsolutions.com