



PhUSE US Connect 2019

Paper AB04

Dashboarding Analytics for Oncology Endpoint Data using SAS

Vidya Muthukumar, Novella Clinical – A Quintiles Company

ABSTRACT

In 2017, of the 946 FDA approvals of clinical trials, 497 were in Oncology (53%). Since oncology trials lead the pack of FDA submissions, more and more CROs are finding newer methods to add analytics into their RFP pipelines to garner new and repeat business. This paper attempts to show how SAS/SAS EG can be used to perform complex analytical methods including survival analysis, mixed models, pattern-mixture, propensity scoring, and sample weighting using oncology endpoint data. Clinical Research Organizations are increasingly pipe-lining analytical offerings in oncology to stand out in the market. Using SAS/SAS EG, one can perform, for example, meta-analysis to assess the association between time-to-disease progression and overall survival in metastatic breast cancer studies and dashboard the analytical results using SAS/SAS EG. In addition, analytics is being used to data mine QOL data from oncology clinical trials in a bid to provide supplemental support to regulatory applications and sponsors.

INTRODUCTION

In the last decade or so, CROs have been adopting innovative techniques to garner a fair share in the oncology trials market. In fact, oncology trials have become more complex than trials in other therapeutic areas. Many of these trials often involve adaptive design and can sometimes include biomarker data as well. As a result, more and more CROs are requiring technically skilled staff with therapeutic experience to work on these complex oncology studies and support their bid defense RFPs. Typically, in oncology studies clinical trial endpoints are of crucial significance. Besides the standard safety data, in the past few years, oncology trials have been collecting additional biomarker data, which can help precisely identify the subpopulation of patients who could benefit from the investigational treatment.

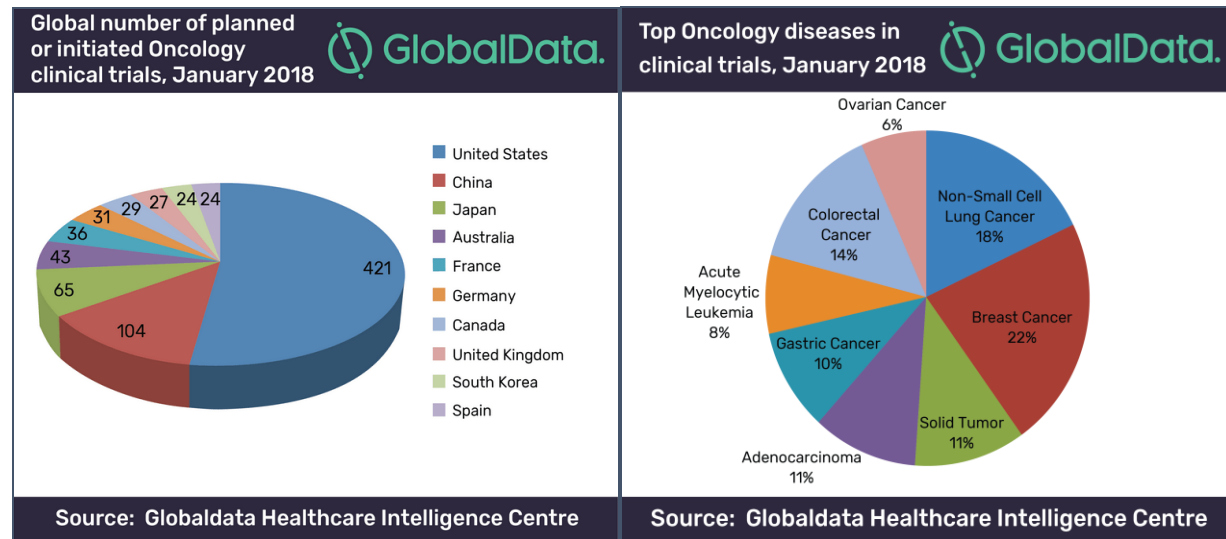
According to the FDA, clinical trial endpoints serve different purposes¹. In conventional oncology drug development, early phase clinical trials evaluate safety and identify evidence of biological drug activity, such as tumor shrinkage. Endpoints for later phase efficacy studies commonly evaluate whether a drug provides a clinical benefit such as prolongation of survival or an improvement in symptoms.

In the study of oncology trials, efficacy analyses have Primary and Secondary endpoints. Primary and secondary endpoints are commonly patients' overall survival (OS) and progression free survival (PFS) and the analysis of overall response rates (ORR). Response refers to the observed change in patient's tumors or lesions. For solid tumors, the standard is provided by RECIST (Response Evaluation Criteria in Solid Tumors) criteria published in 2000 by the European Organization for Research and Treatment of Cancer (EORTC), NCI, and the National Cancer Institute of Canada Clinical Trials Group. Since there hasn't been a cure found for cancer yet, the Quality of Life (QoL) is often one of the secondary endpoints as one can expect some results. CROs periodically present clinical trials data for Interim Analyses and/or for DMC (Data Monitoring Committee) meetings for Oncology studies. Interim Analysis intended to assess treatment effect with respect to efficacy or safety at any time prior to completion of the clinical trial. A DMC is a group of experts external to the study that reviews on a regular basis the key data from an ongoing clinical trial. Many times, CROs use these meetings as a precursor to FDA submissions and/or for new RFPs or bid defense proposals to garner new businesses or get repeat businesses if the trial is progressing smoothly.

For oncology trials, many CROs often periodically submit analytical reports or assessments about the disease progression and use data/visual analytics techniques to perform different kinds of study analyses, either for bid defense or RFPs or to sponsors when providing periodic trials assessments. In oncology trials, the common graphical or analytical Tables/Listings/Figures (TLFs) include Kaplan Meier plots, Forest plots, Waterfall plots, Swimmer plots and Spider plots. Different graphic interpretations of oncology endpoint data have different implications. These graphical analytical interpretations are used more as a tool to assess the impact or significance of the drug or the compound during the term of the oncology trial.

CANCER STATISTICS

In the United States, according to the *American Association for Cancer Research (AACR)*, it is estimated that 1,735,350 new cases of cancer would be diagnosed in 2018 and 609,640 people would have died from some type of cancer disease.² Breast cancer continues to be the top oncology disease in clinical trials with 22% followed closely by non-small cell lung cancer at 18%. According to Globaldata Healthcare, the United States had the largest number of Oncology clinical trials followed by China and Japan in 2018. Also, as of 2018, FDA had approved 63 cancer drugs.³



ONCOLOGY ENDPOINTS

According to *JAMA*⁴, many oncology trials use a set of established methods to compare one cancer treatment with another. Some of these measures define the ability of the treatment to inhibit cancer progression and improve patient outcomes, while others focus on adverse effects, safety, or treatment costs. Clinical trials are designed to identify one factor as the most important goal for the trial, which is its primary endpoint. The earliest research in humans of a new drug or combination, a *phase 1 study*, explores the safety and tolerability of a treatment as drug dose is gradually increased. This process determines *dose-limiting toxic (DLT)* effects, those frequent and severe enough to set the upper limit of dosing for future larger trials. A *phase 2 trial* generally enrolls a few dozen patients treated with one approach or assigned at random (“*randomized*”) to 1 of 2 or more different treatments for their cancer. A phase 3 trial may randomize hundreds of patients to 1 or more novel treatments or to a *control arm* of patients receiving the best treatment available now, referred to as the *standard of care*.

There are several commonly used efficacy end points:

- Objective response rate—the percentage of patients whose cancer shrinks to a specific degree, usually based on imaging assessments
- Progression-free survival (PFS)—the proportion of patients who are alive and whose cancer does not progress on scans over a period from the start of the treatment.
- Overall survival (OS)—the proportion of patients still alive over a period from the start of the treatment.

ANALYTICS FOR ONCOLOGY RFPs

Based on the critical oncology endpoints, many CROs use analytics to garner prospective clinical trials business from RFPs or for generating repeat businesses. Some of the common emerging trends in building analytics into Oncology submissions, RFPs and repeat businesses include the following offerings. Newer web-based and virtual clinical trials are also gaining fast momentum and will see an up climb in the new future.

Data Sources	Analytic Expertise	Deliverables
• Clinical trial studies (Phase I–IV) • Registries and other prospective observational studies • Scientific literature • Claims data • Electronic medical records	• Health-related quality of life (HRQoL) • Q-TWiST (quality-adjusted time without symptoms or toxicities) • Exploratory and post hoc queries • Non-randomized comparisons • Systematic literature reviews • Meta-analysis, including indirect and mixed treatment comparisons	• Protocols • Statistical analysis plans (SAPs) • Tables, listings, and figures (TLFs) • Clinical study reports (CSRs) • Supplementary evidence and support for FDA submissions • Response to regulatory questions • Peer-reviewed publications

ONCOLOGY ENDPOINTS AND ANALYTICS

In SAS/SAS EG and JMP, one can use for example, a SAS Proc called ICLIFETEST to estimate the survival function and test for equality of survival functions by using interval-censored data from a breast cancer study.

In the following sample data that consists of data for 94 subjects from a retrospective study that compares the risks of breast cosmetic deterioration after tumorectomy⁶. There are two treatment groups: patients who receive radiation alone (TRT=RT) and patients who receive radiation plus chemotherapy (TRT=RCT)⁷. Patients are followed up every four to six months, leading to interval-censored observations of deterioration times.

Observations are called censored when the information about their survival time is incomplete⁵; the most commonly encountered censor is right censoring. Suppose patients are followed in a study for 20 weeks. A patient who does not experience the event of interest for the duration of the study is said to be right censored. The survival time for this subject is for the duration of the study. Another example of right censoring is when a subject drops out of the study before the end of the study observation time and did not experience the event. This subject's survival time is said to be censored, since we know that the event of interest did not happen while this subject was under observation. Censoring is an important issue in survival analysis, representing missing data. Censoring that is random and non-informative is usually required to avoid bias in a survival analysis.

In a sample of the 94 observation times, 38 are right-censored and the remaining 56 are censored into intervals of finite length. The following statements create a sample SAS data set named *RT*⁶ for the group that receives radiation alone. The variable *lTime* provides the last follow-up time at which cosmetic deterioration has not occurred for the patient, and the variable *rTime* provides the last follow-up time immediately after the event. Note that for the ICLIFETEST procedure to recognize the observations as right-censored, their right bounds must be represented by missing values in the input data set.

```
data RT;
  input lTime rTime @@;
  trt = ' RT';
  datalines;
.....;
```

The following statements create a SAS data set named *RCT* for patients who receive both radiation and chemotherapy:

```
data RCT;
  input lTime rTime @@;
  trt = 'RCT';
  datalines;
.....;
```

The following statements combine the data sets *RT* and *RCT* into a single data set called *BCS* that is to be analyzed by the ICLIFETEST procedure:

```
data BCS;
  set RT RCT;
run;
```

Suppose you want an insight into the incidence rate of cosmetic deterioration over time after the two treatments. From the perspective of survival analysis, the tasks are to estimate the survival probabilities for the two treatment groups and to test for a systematic difference between the groups.

The following statements invoke the ICLIFETEST procedure to estimate the survival functions for both treatment groups:

```
ods graphics on;
proc iclifetest plots=(survival logsurv) data=BCS impute(seed=1234);
  strata trt;
  time (lTime, rTime);
run;
```

In the TIME statement, the variables that represent the interval boundaries, *lTime* and *rTime*, are noted in the example below. Because the treatment indicator variable, *Trt*, is specified in the STRATA statement, PROC ICLIFETEST conducts the analysis separately for each treatment group. When you specify the keywords SURVIVAL and LOGSURV in the PLOTS= option, the procedure plots the estimated survival functions and the negative of the log transformations of the estimates. You can specify an integer seed for the random number generator that is used in

creating imputed data sets for calculating standard errors of the survival estimates. If the SEED= option is not specified, a random seed is obtained from the computer's clock settings.

The ICLIFESTProcedure

Data and Methods Information	
Data Set	WORK.BCS
Left Boundary	ITime
Right Boundary	rTime
Estimation Method	EMICM
Number of Observations Read	94
Number of Observations Used	94
Number of Imputations for Standard Errors	1000
Imputation Seed	1234

Page Break

The ICLIFESTProcedure

Stratum 1: trt = RT

Nonparametric Survival Estimates				
Time Interval		Probability Estimate		Imputation Standard Error
		Failure	Survival	
0	4	0.0000	1.0000	0.0000
5	6	0.0463	0.9537	0.0354
7	7	0.0797	0.9203	0.0458
8	11	0.1684	0.8316	0.0580
12	24	0.2391	0.7609	0.0629
25	33	0.3318	0.6682	0.0706
34	38	0.4136	0.5864	0.0739
40	46	0.5344	0.4656	0.0758
48	Inf	1.0000	0.0000	0.0000

NOTE: The estimation algorithm converged.

Quartile Estimates

Percentile	Estimate	95% Confidence Interval	
		Transform	Lower Upper
75	36	LOGLOG	25 36
50	20	LOGLOG	19 25
25	17	LOGLOG	12 19

Page Break

The ICLIFESTProcedure

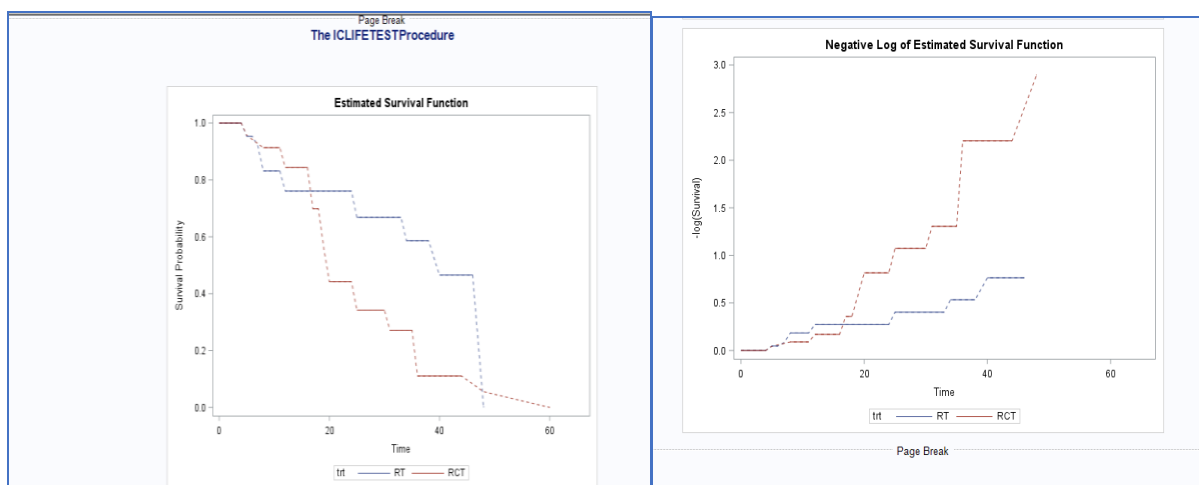
Stratum 2: trt = RCT

Nonparametric Survival Estimates				
Time Interval		Probability Estimate		Imputation Standard Error
		Failure	Survival	
0	4	0.0000	1.0000	0.0000
5	5	0.0433	0.9567	0.0332
8	11	0.0866	0.9134	0.0413
12	16	0.1558	0.8442	0.0567
17	18	0.3012	0.6988	0.0764
19	19	0.4423	0.5577	0.0835
20	24	0.5680	0.4420	0.0771
25	30	0.6579	0.3421	0.0740
31	35	0.7288	0.2712	0.0674
36	44	0.8896	0.1104	0.0482
48	48	0.9448	0.0552	0.0352
60	Inf	1.0000	0.0000	0.0000

NOTE: The estimation algorithm converged.

Quartile Estimates						
Percentile	Estimate	95% Confidence Interval				
		Transform	Lower	Upper		
75	36	LOGLOG	25	36		
50	20	LOGLOG	19	25		
25	17	LOGLOG	12	19		

Number of Censored and Uncensored Values						
Stratum	ID	trt	Type of Censoring			Uncensored
			Total	Left	Interval	
1	RT		46	3 (6.5%)	18 (39.1%)	25 (54.3%)
2	RCT		48	2 (4.2%)	33 (68.8%)	13 (27.1%)
Total			94	5 (5.3%)	51 (54.3%)	38 (40.4%)





While analytics is being increasingly used to make newer offerings via RFPs or to generate repeat businesses for many CROs, the results from the visual analytical findings like the above sample data is often summarized in an analytical dashboard and presented to prospective vendors and bidders when competing for trial awards such as complex oncology trials. The analytics can be endless. But it is always crucial to dish out the key findings from complex oncology studies and present the results in an "easy-to comprehend" manner to garner prospective business.

Another example commonly used in analytics of oncology trials is the use of Butterfly Plots using SGPLOT. The following is sample data from *the American Cancer Society* and used in SAS/STAT⁶ to generate the graph.

```
data work.cancer;
infile datalines;
input cause $ 1-20 mcases fcases mdeaths fdeaths;
deaths=mdeaths + fdeaths;
mcases= -1 * mcases;
mdeaths= -1 * mdeaths;
datalines;

Lung Cancer      114760 98620 89510 70880
Colorectal Cancer 55290 57050 26000 26180
Breast Cancer    2030 178480 450 40460
Pancreatic Cancer 18830 18340 16840 16530
Prostate Cancer  218890 0 27050 0
Leukemia         24800 19440 12320 9470
Lymphoma         38670 32710 10370 9360
Liver Cancer     13650 5510 11280 5500
Ovarian Cancer   0 22430 0 15280
Esophageal Cancer 12130 3430 10900 3040
Bladder Cancer   50040 17120 9630 4120
Kidney Cancer    31590 19600 8080 4810
;
run;

proc sort data=cancer;
  by descending deaths;
run;

proc format;
  picture positive
    low-<0='000,000'
    0<=high='000,000';
run;

title 'Leading Causes of US Cancer Deaths in 2007';
footnote justify=left italic 'Source: American Cancer Society';

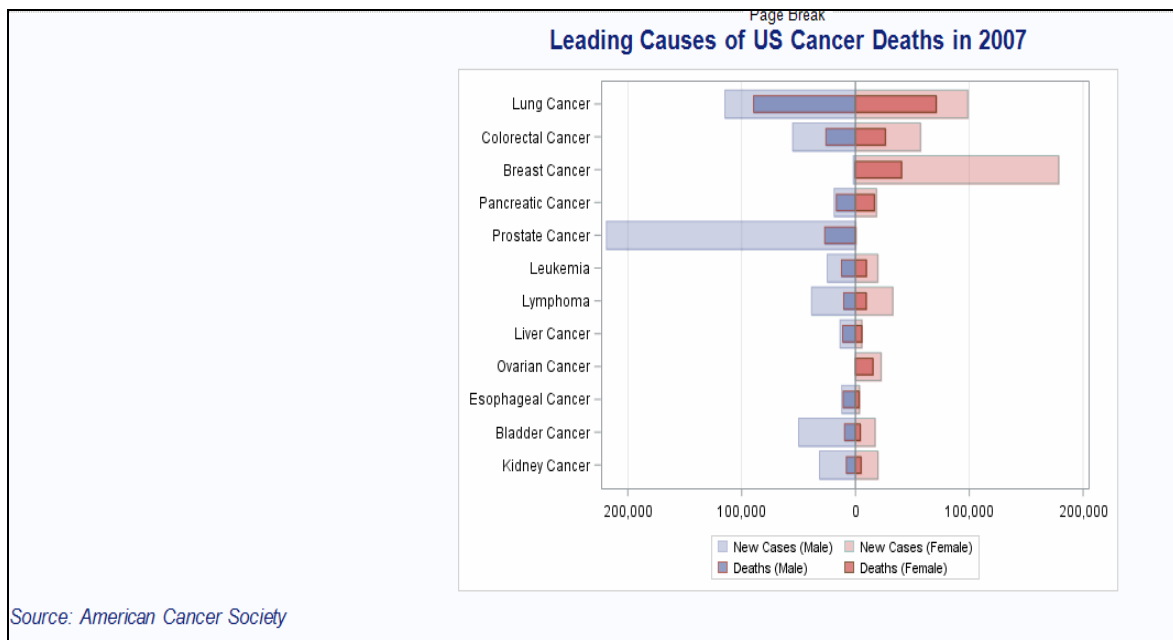
ods listing close;
ods html file='CancerDeaths' path='.';
ods graphics / reset width=600px height=400px imagename='Cancerdeaths' imagefmt=gif;

proc sgplot data=cancer;
  format mcases mdeaths fcases fdeaths positive.;
  hbar cause / response=mcases
    fillattrs=graphdata1 transparency=.65
    legendlabel="New Cases (Male)" name="mcases" ;
  hbar cause / response=mdeaths barwidth=.5
    fillattrs=graphdata1 transparency=.25
```

```

legendlabel="Deaths (Male)" name="mdeaths" ;
hbar cause / response=fcases
fillattrs=graphdata2 transparency=.65
legendlabel="New Cases (Female)" name="fcases";
hbar cause / response=fdeaths barwidth=.5
fillattrs=graphdata2 transparency=.25
legendlabel="Deaths (Female)" name="fdeaths";
keylegend "mcases" "fcases" "mdeaths" "fdeaths" / across=2;
axis display=(nolabel) grid;
yaxis display=(nolabel) discreteorder=data;
run;
ods html close;
ods listing;

```



Depending on the purpose of the analyses or the use of analytics collected during the oncology trials, different visual/data analytics can be performed using SAS/SAS EG, JMP or SAS Visual Analytics. The platforms are many, but the key is to accurately perform the analytics and to have some form of standardization when using analytics for evaluating oncology study endpoints. This would depend on the intent of the business decision, whether it is for RFPs or for generating repeat business or for performing pre-submission evaluations to audit the study progression, especially for oncology trials.



CONCLUSION

While the use of Analytics in oncology trials is increasingly gaining ground, the published Consolidated Standards of Reporting Trials (CONSORT) best-practice guidelines encourage the reporting of clearly defined primary and secondary outcome measures. While Overall Survival is the standard of endpoints, but as increasing numbers of treatments become available for different types of cancer, and as more cohorts of patients are included; this requires a longer follow-up period and increases the cost of clinical trials. Thus, tumor-centered clinical endpoints that can be assessed earlier and used as surrogates for overall survival are increasingly being studied, but most of them currently lack standardized definitions to enable cross comparison of results among different clinical trials and they have not been validated as surrogate endpoints. In addition, the variability of their definition can strongly impact the trial's conclusions by affecting both statistical power and estimation. In fact, QoL has slowly become a useful surrogate endpoint for trials since this endpoint can ensure treatment benefit to some extent from both the patient and public health points of view. Methodological research should be pursued to develop standard outcome definitions for use in cancer clinical trials and to define a standardized longitudinal analysis of QoL data.

REFERENCES

1. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry. FDA document. <https://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf>
2. <https://www.cancerprogressreport.org>.
3. <https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm279174.htm>
4. <https://jamanetwork.com/data/journals/oncology/937694/cpg180005fa.png>
5. Finkelstein, D. M. (1986). "A Proportional Hazards Model for Interval-Censored Failure Time Data." *Biometrics* 42:845–854.
6. <https://documentation.sas.com>
7. Kleinbaum D.G., Survival Analysis, a self-learning text, Springer-Verlag, 1996.

CONTACT INFORMATION

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

Author Name: Vidya Muthukumar
Company: Novella Clinical – A Quintiles Company
Address: 1700 Perimeter Park Drive
City / Postcode: Morrisville, NC 27560
Work Phone: 984-212-5831
Email: vidya.muthukumar@novellaclinical.com
Web: www.novellaclinical.com

Brand and product names are trademarks of their respective companies.