PharmaSUG 2022 - Paper AD-036

Reconstruction of survival curves: An approach of comparing study drug curve with the standard of care digital survival curve

Girish Kankipati and Jai Deep Mittapalli, Seagen Inc., Bothell WA

ABSTRACT

Survival analysis corresponds to a set of statistical approaches used to investigate the time it takes for an event of interest to occur. Most of the survival procedures involve KM plots, log-rank tests, etc. Survival analysis focuses on two important aspects: occurrence or non-occurrence of an event, and follow-up time. We are often interested in assessing whether there are differences in survival (or cumulative incidence of event) among different treatments. For instance, in first-in-human (FIH) clinical trials we want to compare survival curves between the investigational and comparative drugs on market (or the standard of care). For exploratory purposes, there is often an interest in seeing study vs. comparison drug survival curves on the same graph. To plot the survival curve for standard of care, we may need to construct the data from a digital image file in .png or .jpg format. How do we do it?

To address these challenges, this paper introduces an innovative approach for reconstruction of data and comparison of survival curves. It will cover how to extract data points from survival curve digital images, using the trialdesign.org site as an example. The data points extracted from the image are saved to an Excel file which is then converted into a SAS® data set and combined with study drug survival data to generate the plot.

INTRODUCTION

In survival data analysis, we need to estimate the overall survival of a particular population under study which is also called Time to Event analysis. Clinical trials often collect and calculate time to event data for patients. The data can be used to compare two or more groups for the rate at which an event of interest occurs. In some special cases, study teams often request programmers to compare the investigational drugs with the comparative drug. In this scenario, it's difficult to compare the data from a published Portable Network Graphics (*.png) or JPEG (*.jpg) files. To perform such a comparison, this paper presents a solution on extraction of data points from a survival curve digital image. We will discuss the process of extracting the data points of interest from the previously published standard KM plots into an Excel spreadsheet and convert it into a SAS® data set. That way, these data points can be read into the figure program.

READING PNG OR JPEG FILES INTO SAS

Trialdesign.org is the website used to convert the published data into data points. The initial step is to upload the image into the Trial design website; subsequent steps to extract the data are listed below.

Step1: Set up the maximum survival time on the x-axis, for example 40.
Step2: On the X axis, click at the origin 0
Step3: On the Y axis, click on the point where the survival probability is 0
Step4: On the Y axis, click on the point where the survival probability is the maximum
Step5: Extract the data points of interest in the graph and then download the Excel file to convert into a SAS data set
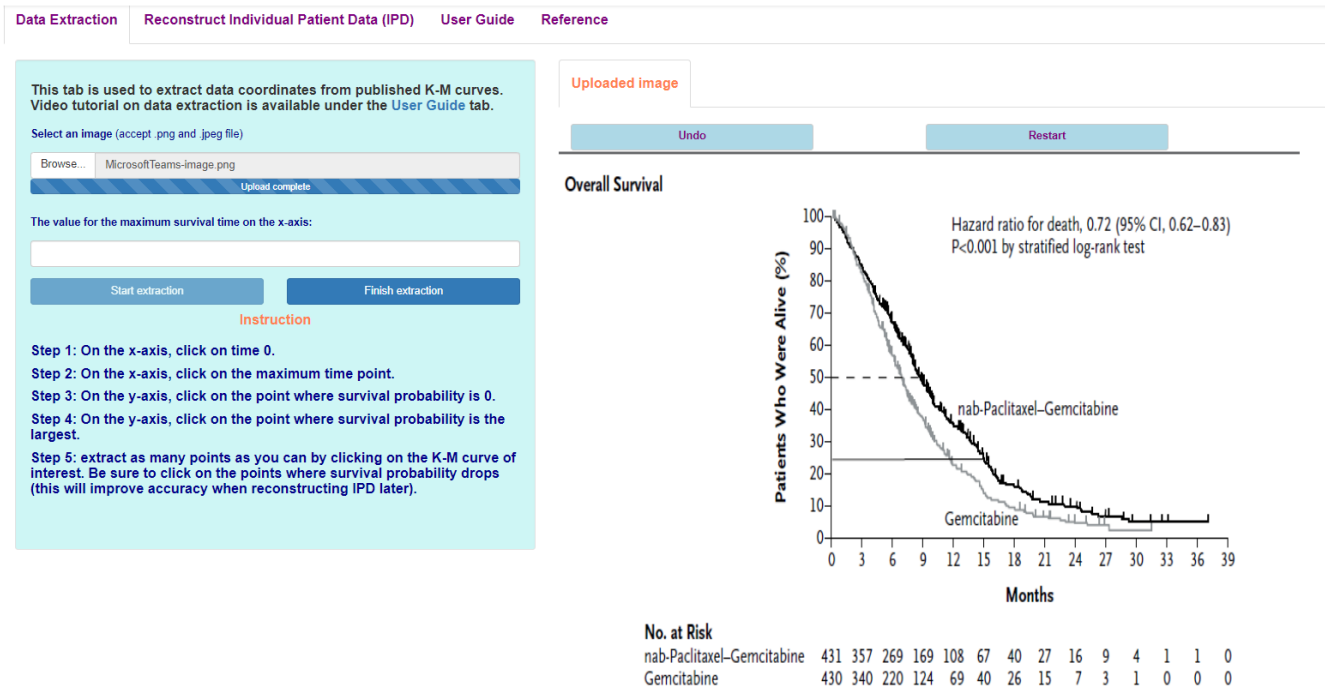
EXTRACTION OF DATA POINTS:

Figure 1: Extraction of data points using trialdesign.org

After the extraction is done, click "Finish extraction" so that all the points will get extracted into an Excel file. Below is the example from the data used in this paper. The data contains timepoints and survival rate of the curve.

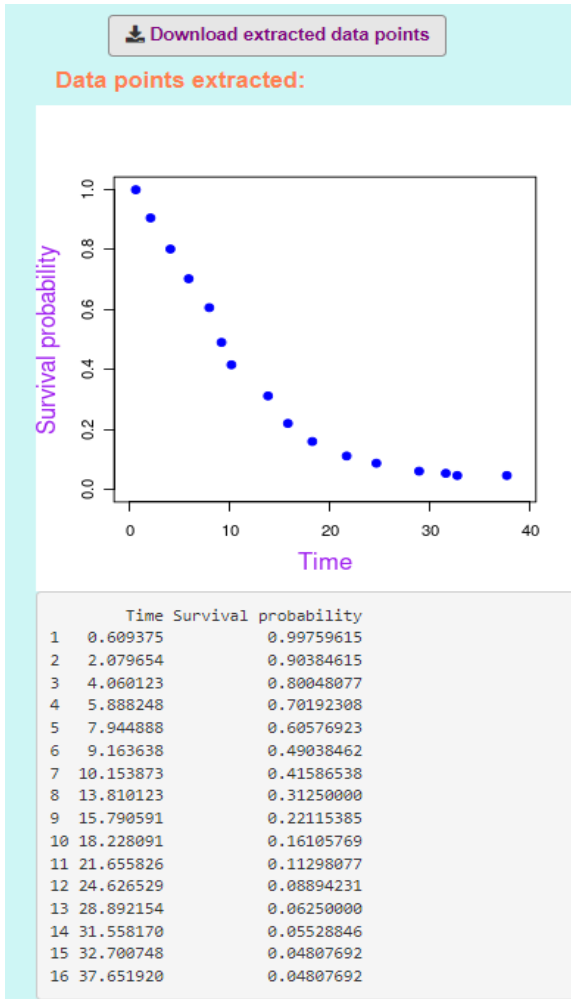| Time	Survival
0.609375	0.997596154
2.079653978	0.903846154
4.060122728	0.800480769
5.888247728	0.701923077
7.944888353	0.605769231
9.163638353	0.490384615
10.15387273	0.415865385
13.81012273	0.3125
15.79059148	0.221153846
18.22809148	0.161057692
21.65582585	0.112980769
24.62652898	0.088942308
28.89215398	0.0625
31.5581696	0.055288462
32.70074773	0.048076923
37.6519196	0.048076923

Table 1: Excel table

Figure 2: Extraction of data points from the digital curve

In the above **Figure 2**, up to 39 months of survival data are extracted from the digital .png file. The data contains time and survival probability, and these two values are sufficient to plot the values in a KM figure. Using SAS®, we can import the Excel file and convert it into a SAS® data set. The figure program will read the extracted data for the competitive drug and the data from the ADaM ADTTE data set for the investigational drug. These data are set together, and we treat them as two different treatment groups. In the program, we used PROC TEMPLATE and SGRENDER to plot the survival curves. An R Shiny app is another method to plot KM curves for individual patient data (IPD).

```sas
proc format;
    value trtn
        1="Published Patient Data"
        2="Investigational Drug";
    value cen
        1="Published Patient Data"
        2="Investigational Drug";
run;

/*Set the Published data points and Study Treatment points, both data sets must have common variables such as
AVAL (Survival) and Time */

data surv;
    set out.points_extracted (in=a)  out.treatment (in=b);

    if a then trt01pn=1;

    if b then trt01pn=2;
run;

proc sort data=surv;
    by trt01pn subject;
run;

ods listing close;
ods output Quartiles=col3;
ods output CensoredSummary=cen;
ods output ProductLimitEstimates=risk;

proc lifetest data=surv method=pl outsurv=km;
    strata trt01pn /test=logrank;
    time aval*cnsr(1);
run;

ods output ParameterEstimates=col2;

proc phreg data=surv;
    class trt01pn;
    model aval*cnsr(1)= trt01pn / rl;
run;

data all;
    set km;
    format trt01pn trtn. trt2pn cen.;
    trt2pn= trt01pn;
    keep aval survival trt01pn trt2pn _CENSOR_;
run;

/*Assign the labels for stats*/
%let c1r1=;
%let c1r2=;
%let c1r0=;
%let c2r0=;
%let c3r0=;
%let c4r0=Median;
%let c5r0=;
%let c6r0=;
```

```sas
%let c1r01=;
%let c2r01=N;
%let c3r01=Events;
%let c4r01=(Months);
%let c5r01=HR (95% CI);
%let c6r01=p-value*;

data outmed;
   * Median;
   set col3(where=(percent=50));
      length estimate_ $10;
   if estimate ne . then
      estimate_=strip(put(round(estimate,.01), best.));
   else
      estimate_='-';
   call symputx(choosec(trt01pn,'c4r1','c4r2'),estimate_);
run;

data _null_;
   * N and Events;
   set cen;

   if TRT01PN=2 then
      do;
         call symputx ('c2r2', put(total, 8.));
         call symputx ('c3r2', put(failed,8.));
      end;
run;

/**/
proc template;
   define style sty_km;
      parent=rtftnr10;
      style graphfonts from graphfonts /
         'GraphDataFont' =("times new roman", 10pt)
         'GraphDataValue'=("times new roman", 10pt)
         'GraphLabelFont'=('times new roman', 10pt)
         'GraphTitleFont'=('Arial',11pt)
         'GraphFootnoteFont'=('times new roman',10pt);

      *-----------------------------------------------------*
       | **Analysis Specific**
      *-----------------------------------------------------*;
      style graphdata1 from graphdata1 /color = DarkGreen contrastcolor = DarkGreen linestyle=1;
      style graphdata2 from graphdata2 /color = DarkBlue contrastcolor = DarkBlue linestyle=1;

      style graphdata3 from graphdata3 /markersymbol = "circlefilled"   contrastcolor = DarkGreen;
      style graphdata4 from graphdata4 /markersymbol = "trianglefilled"  contrastcolor = DarkBlue;
   end;
run;

proc template;
   define statgraph km_ar;
   mvar  c1r0 c2r0 c3r0 c4r0  c1r01 c2r01 c3r01 c4r01  c1r1 c1r2 c2r1 c2r2 c3r1 c3r2 c4r1 c4r2;
   begingraph / designwidth=9in designheight=4.5in border=false;
   entrytitle ' ';
   layout lattice /columndatarange=union rowweights=(0.85 .05 .1);
```

```sas
*_____*
| Survival plot
*_____*;
layout overlay / yaxisopts=(label='Survival' linearopts=(viewmin=0 viewmax=1)) cycleattrs=true
xaxisopts=(label="Time (Months)"  offsetmin=0.01 offsetmax=0.01
linearopts=(viewmin=0 viewmax =42
tickvaluelist=(3 6 9 12 15 18 21 24 27 30 33 36 39 42)));
stepplot x=aval y=survival /  group=TRT01PN name='km' lineattrs=(thickness=2pt);
scatterplot x=eval(choosen(_censor_+1,.,aval)) y=survival/
   group=trt2pn name='cen' legendlabel="Censored Subjects";

*** Pad Right & Top in following statement determines location of group labels for group legend. ***;
discretelegend 'km'  / location=inside halign=right pad=(right=150 top=28)
   valign=top across=1 border=false valueattrs=(size=8pt);

*_____*
| Embedded table in survival plot
*_____*;
layout gridded /valign=top halign=right border=false
   opaque=false columns=4 order=rowmajor;
   entry halign=center c1r0 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c2r0 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c3r0 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c4r0 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c1r01 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c2r01 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c3r01 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c4r01 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c1r1 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c2r1 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c3r1 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c4r1 / textattrs=(size=8pt color=DarkGreen);
   entry halign=center c1r2 / textattrs=(size=8pt color=DarkBlue);
   entry halign=center c2r2 / textattrs=(size=8pt color=DarkBlue);
   entry halign=center c3r2 / textattrs=(size=8pt color=DarkBlue);
   entry halign=center c4r2 / textattrs=(size=8pt color=DarkBlue);
endlayout;
endlayout;
endlayout;
endgraph;
end;
run;

title1 j=l "Combining Published Patient Data with Study Drug Survival Data";
ods _all_ close;
ods proclabel=' ';
ods graphics / imagefmt=emf;
options orientation=landscape;
ods rtf file="&outputname" path="&outdir" style=sty_km  nogfootnote nogtitle   bodytitle toc_data  headery=720
footery=720;

proc sgrender data=all template=km_ar objectlabel="&hidetitle";
run;

ods rtf close;
ods listing;
```
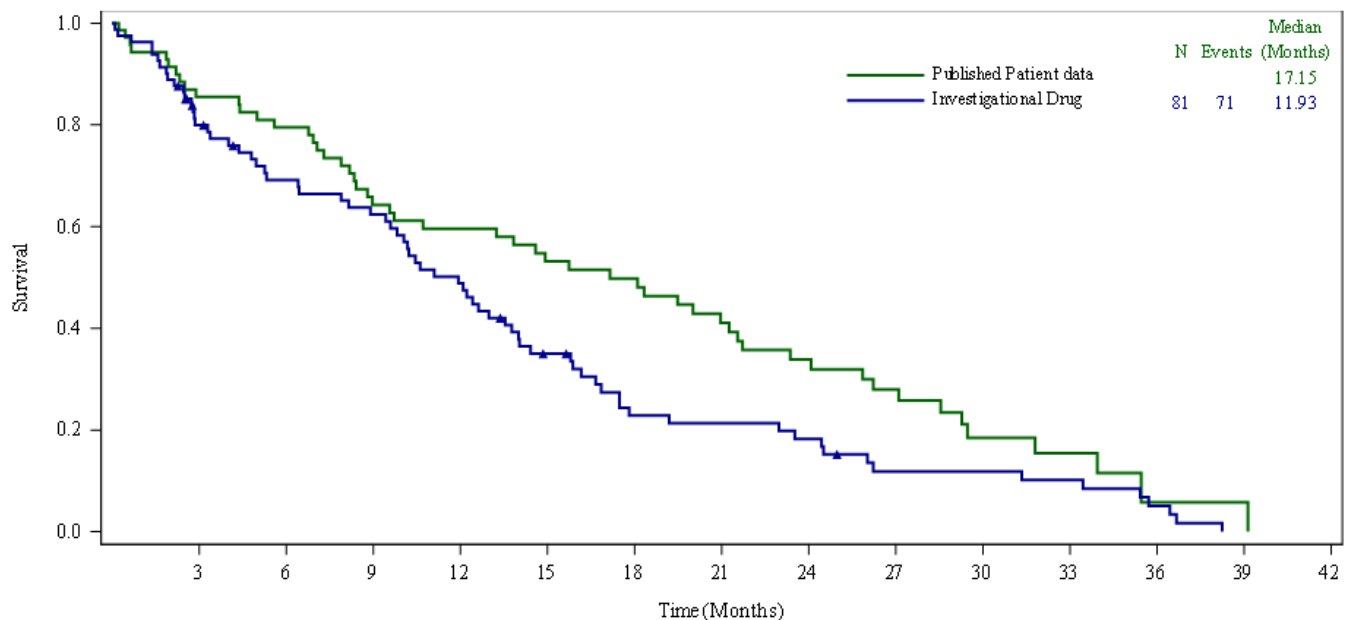
COMBINING THE CURVES

Figure 3: Extraction of data points from the digital curve

In the above figure, survival curves were shown in different colors against time in months on the X-axis and survival rate on the Y axis. In this example, N and Events values are blank for the published patient data. This is because the censor values were not extracted in the process, only the time and survival values were saved. Other values like P-values and Hazard ratios were not calculated for the same reason. On the other hand, since the investigational study drug is being examined in an ongoing study, in that data we are able to find event information and populate that in the legend for Event and N.

CONCLUSION

This paper introduces a new method to compare published standard of care .png data to ongoing clinical trials on a study drug under investigation. This method helps programmers, statisticians, and clinicians to display and compare both curves on the same plot.

REFERENCES

Daniel D. Von Hoff, M. T. (n.d.). *The New England Journal of Medicine.* Retrieved from
https://www.nejm.org/doi/full/10.1056/NEJMoa1304369

Liu, N. Z. (n.d.). *Trial Design*. Retrieved from https://trialdesign.org/one-page-shell.html#IPDfromKM

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Girish Kankipati
Seagen Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
gkankipati@seagen.com

Jai Deep Mittapalli
Seagen Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
jamittapalli@seagen.com