PharmaSUG China 2021 - Paper SR-041

Number of Events Estimation using R step by step

ChunYan He, Ark Biosciences Inc.;
LiYan Zhao, Changchun University of Technology

ABSTRACT

Estimated enrollment time and follow-up time are used for calculating sample size in oncology clinical trial when time to event is considered as primary endpoint. However, enrollment time and follow-up time might be different from what we specified in protocol during clinical trial conduction due to kind of reasons. If they are quite different in reality, analysis timepoint we estimated will not be accurate any more. In the case, we usually want to know when the analysis timepoint will occur or how many events will happen by specified follow-up time (e.g., 6 month). We will demonstrate what information should be considered when addressing this question and simulations using R step by step are introduced in this paper.

INTRODUCTION

Below figure pictorially describes the survival experience of subjects in a trial; from the start of treatment some subjects progress to the event (here death) or are censored. Actually, this figure is somewhat artificial as it assumes everyone arrives at the same time simultaneously and then is followed up for observation of whether the event has occurred.

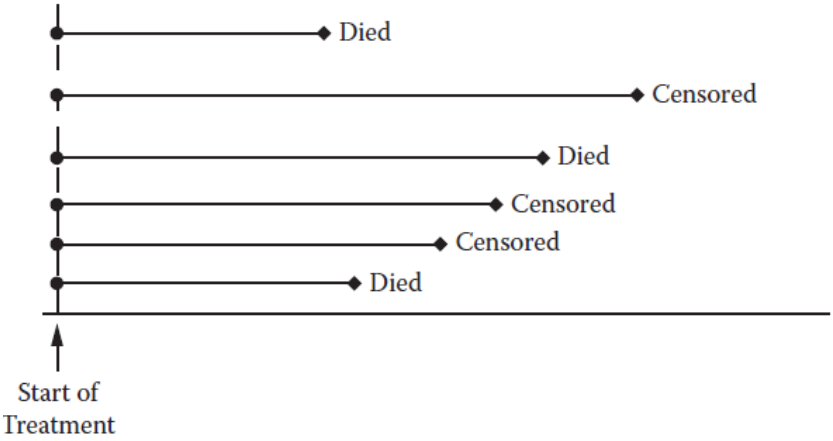


Figure 1. The time course of the trial

In actuality Figure 2 more accurately represents the time course of a trial because following the study start subjects are recruited for a period of time. This recruitment (accrual) period ends after a period of time, and then at a fixed point after this accrual period time the study ends, and subjects are analyzed. Hence all subjects may be in for a minimum period of time, but the actual period of time subjects may have been in the trial may vary quite markedly. Another complicating feature is that we have a study end at which we need to undertake a statistical analysis. Of course, if we waited long enough all subjects would reach the survival endpoint in particular, but at a given time point we perform a statistical analysis. So, recruitment time and follow-up time are critical information for sample size estimation.

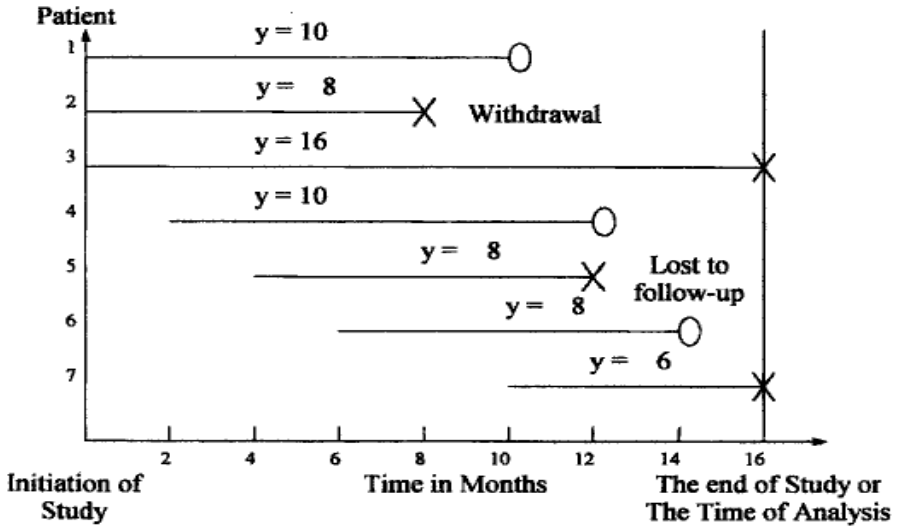


Figure 2.The time course of the trial

With respect to sample size calculations in oncology clinical trials when time to event as primary endpoint, it is the number of events that is of important, and a sufficient number of subjects should be recruited to ensure a sufficient number of events. Therefore, analysis timepoint is established according to events occurred rather than number of subjects enrolled.

Estimated enrollment time and follow-up time are used for calculating sample size in oncology clinical trial when time to event as primary endpoint. However, enrollment time might be different from what we specified in protocol during clinical trial conduction due to kind of reasons. Short enrollment time will need long term follow-up in order to achieve sufficient events. While long-term enrollment time with shorten follow-up time can achieve required events. Therefore, if recruitment time in reality is quite different from previous established value, analysis timepoint we estimated previously will not be accurate any more. In the case, study team usually want to know when the analysis timepoint will occur or how many events will happen by specified follow-up time (e.g., 6 month). In below sections, we will demonstrate simulation steps by an example.

ASSUMPTIONS FOR SIMULATION

Before addressing this question, we need to discuss with team to well know the actual recruitment rate. Sample size, estimated median survival time, and drop-out rate specified in protocol are also needed. Here we will use overall survival (OS) as our single-arm sample study endpoint with median OS is 4.8 months. Suppose we will recruit 48 subjects and actual enrollment time is 6 months. We assume OS follow exponential distribution. According exponential function, lambda is equal to Ln2/mOS. For simplicity, we assume survival time are right censored and enrollment time follows uniform distribution.

Items	Values
Study Design	Single-arm
Sample size	48
Median OS	4.8 months
Enrollment time	6 months
Enrollment time distribution	Uniform distribution [0, 6]

Drop-out %	5%
Survival time distribution	Exponential distribution
Lamda in survival time exponential distribution	$\frac{\mathrm{Ln}2}{4.8}$
Survival time censoring type	Right censored
Censoring time distribution	Exponential distribution
Estimation method of S(t)	Kaplan-Meier product-limit estimator
Number of Simulation	100

Table 1. Simulation Assumptions

Question 1: When 70% subjects will have events (death)?

Solution to Q1：To estimate 70% quantiles

Question 2: How many percent of subjects will have events by 6 months?

Solution to Q2：To estimate number of subjects with events by 6 months，%=$\frac{n_j}{n}$

SIMULATION AND RESULT

SIMULATION

Simulation steps

In order to solve above two problems, R Studio is used for simulation research. The simulation process is divided into following six steps.

1. Simulate survival time according to survival distribution parameter setting.
2. Simulate enrollment time according to enrollment distribution parameter setting.
3. Simulate censoring time according to the type of censoring, censoring distribution parameter setting.
4. Produce survival data according to simulated survival time, enrollment time and censoring time created in Step 1~3.
5. Estimate 70% Quantiles for each simulation trial using Kaplan-Meier methods and calculate means of 70% quantiles from simulation trials.
6. Estimate number and percent of subjects with events by specified timepoint for each simulation trial. Calculate means of percentage from simulations trials.

Simulation program

Install and load the packages

```
install.packages("survival")
library("survival")
```

Initialize sample data

```
set.seed(99);
n=48;                                   ### sample number
cycle=100;                              ### cycle-index
storVar=matrix(0, cycle,1);             ### the storage variable for the result
```

```
EnrollT0=6;                                  ### enrollment time
lamda0=4.8;                                  ### Median OS
lamda=log(2)/lamda0;                         ###survival time's parameter
```

Start of Circulation

```
for (i in 1:cycle)
{
```

Simulate OS

```
    OS=rexp(n,lamda);
```

Simulate enrollment time

```
    EnrollT=runif(n,0, EnrollT0);
```

Simulate censoring time

```
    c1_seq = seq(0.001,15,0.01);
    c1_iter = NULL;
    c1_est = NULL;
    noncensor_rate = 0.95;
    for(iter in 1:100)
    {
        for(w in 1:length(c1_seq))
        {
            C=rexp(n,c1_seq[w]);
            v=1*(OS <= C);
            if(abs(sum(v/n)-noncensor_rate)<0.01)
              {c1_iter[w]=c1_seq[w];}
            else {c1_iter[w]=0;}
        }
    c1_est[iter]=mean(c1_iter[which(c1_iter!=0)]);
    }
    c1=mean(c1_est,na.rm=TRUE);
    C=rexp(n,c1);
```

Calculate event indicator variable

```
    event=1*(OS<=C);
```

Calculate event observation time

```
    T=pmin(OS,C);
```

Calculate calender time

```
    calender_time =T+EnrollT-EnrollT0;
```

Question 1: Estimate quantile using Kaplan-Meier method

```
    data1=data.frame(calender_time,event)
    fit=survfit(Surv(calender_time,event)~1,data=data1,conf.type="log-log");
    qua=quantile(fit,probs=0.7);
    storVar[i,]= qua$quantile;
}
```

Calculate means of quantiles from above simulation

```
qua_70 =colMeans(storVar);
```

Question 2 : Calculate the frequency of events occurring by 6 months

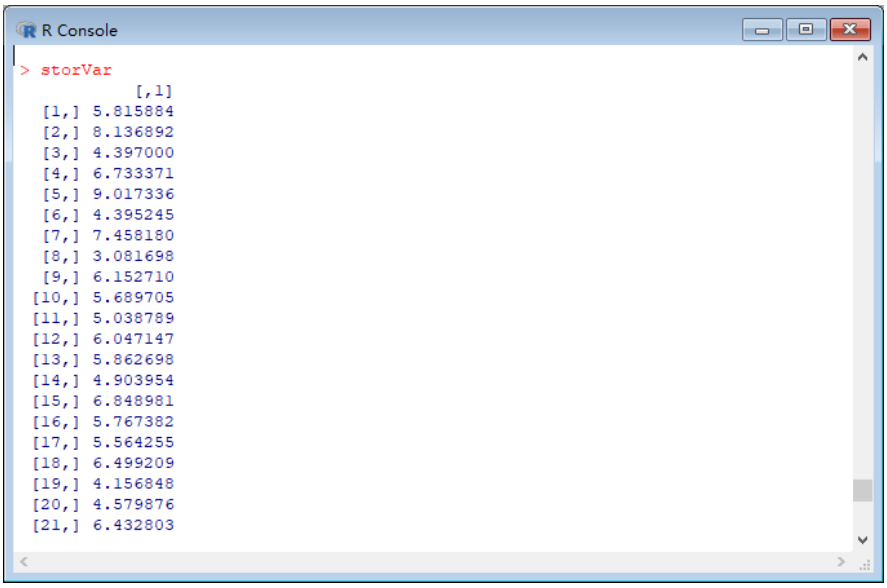
```
    data1=data.frame(calender_time,event)
    data2=data1[which(data1$ event ==1),]
    storVar[i]=length(which(data2$ calender_time <=6))/n
}
```

Calculate means of frequency of events from above simulation

```
nosub_m6=colMeans(storVar);
```

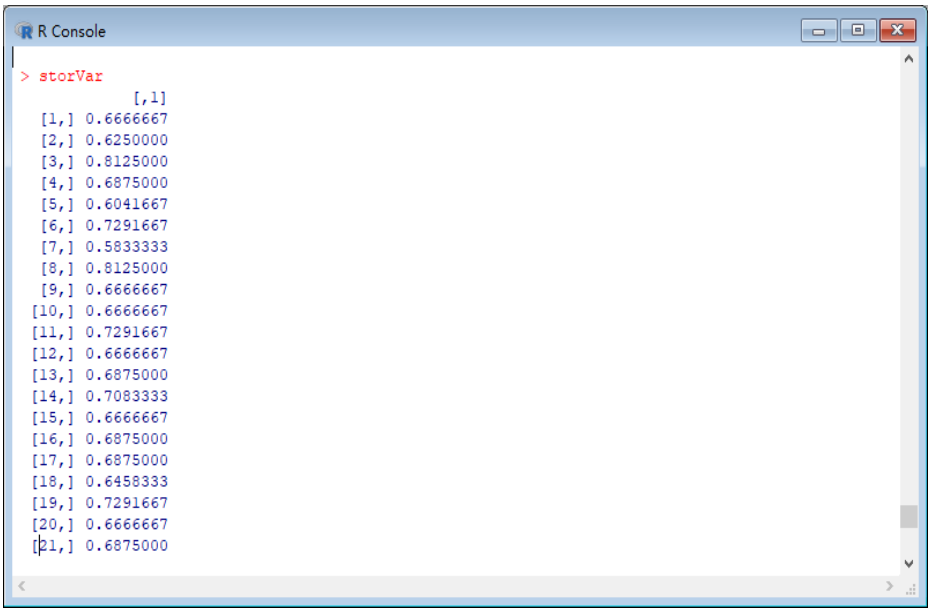
RESULT

Through the simulation, we calculate each 70% quantile using Kaplan-Meier methods from 1000 simulation trials for Question 1 (Screenshot 1). Due to the randomness of the simulation data, the results of 1000 cycles need to be averaged, and the final result is 5.55 months. So, we estimate there will be 70% subjects with events by 5.55 months after all the participants are enrolled.



Screenshot 1. Partial results of 100 cycles of Q1

Similarly, Screenshot 2 also shows part of the results from 1000 simulation trials for Question 2. The average result of above simulation results is 70.22%. That means 70.22% of subjects will have events by 6 months by estimation.



Screenshot 2. Partial results of 100 cycles of Q2

CONCLUSION

Most of assumptions in this paper are with certain limitations. For OS endpoint, right censoring mechanism are applicable. However, in reality, we often use the surrogate endpoint, such as progression free survival (PFS) as primary endpoint. In this case, right censoring is not appropriate anymore and

interval censoring needs to be applied instead. Besides, recruitment rate in real life is usually slow at first, then increasing gradually, and then stable finally. So uniform enrollment rate is somewhat too idealistic. In above example, we assume censoring are all caused by drop-out. However, there are all kinds of censoring reasons in real study. So, we just provide a preliminary idea for number of events estimations for reference.

REFERENCES

Chow, S. C., Shao, J., Wang, H., & Lokhnygina, Y. 2017. *Sample size calculations in clinical research.* Chapman and Hall/CRC.

Qian, J., Liu, G., & Zhou, Y. M., 2013. “Simulation of survival data generation under different deletion” *ratio.Journal of Mathematical Medicine*,26(6), 644-646

CONTACT INFORMATION

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

Name: Chunyan He
E-mail: chunchun005bb@163.com

Liyan Zhao
zhaoliyan10166@163.com