MWSUG 2018 - Paper HS-083

A Macro to Calculate Sample Size for Studies Using the Proportional Time Assumption

Brian Mosier, University of Kansas Medical Center, Kansas City, KS
John Keighley, University of Kansas Medical Center, Kansas City, KS
Milind Phadnis, University of Kansas Medical Center, Kansas City, KS

ABSTRACT

Sample size calculations for time-to-event outcomes are done mostly based on the assumption of proportional hazards or of exponentially distributed survival times. These assumptions are not appropriate for all scenarios and should not be implemented if the assumptions are not met. Phadnis et al[1] introduces an alternative method using the assumption of proportional time by using the generalized gamma ratio distribution to calculate sample size. We developed a macro to calculate sample size needed for studies using the proportional time assumption for a given value of power in an efficient way. The macro automates the method from the paper to simulate survival data for two treatment arms with the test statistic following a generalized gamma ratio distribution. It then utilizes the bisection method in order to find the appropriate sample size needed for power, input by the user along with additional parameters. We have implemented various features in the macro, allowing for one or two-sided tests and an option that graphs the power function (additional features). This macro is a tool that statisticians can use to make sample size calculations for studies using the proportional time assumption when some form of historical information is available from a prior study.

INTRODUCTION

Sample size calculation is an integral part of planning a time-to-event study. To do this properly, statisticians must have the necessary tools to make these calculations. With novel study designs come new challenges for planning these studies. The proportional time assumption assumes that the amount of time it takes for $p\%$ of the population to experience an event is proportional between the treatment and control groups Phadnis et al[1]. For example, with a proportionality time ratio of 2, it will take the treatment group twice as long as the control group for $p\%$ of the population to experience an event. There is currently no software that offers sample size calculation for time-to-event studies using this assumption. This paper will discuss a SAS® macro that uses simulation from the generalized gamma ratio distribution to calculate sample size for studies using this assumption.

We will look at the features of the macro, as well as an example.

MACRO FEATURES

Full code is available via the link at the end of paper.

Oftentimes investigators are unsure of model parameters when planning a study. This can be remedied with data from prior studies. This macro allows users to input data to get these estimates. Additionally, the macro can use data from a prior study to estimate an overall event rate for the study, based on accrual and follow-up times, as well as the given effect size defined in terms of a proportionality time ratio. Estimating the event rate is important for sample size calculation because it adjusts the sample size estimates. If these values are known in advance, they can simply be input into the macro call.

Additionally, other parameters may be input such as power, significance level, loss to follow-up rate, whether you want to test a one or two-tailed test, allocation ratio, and effect size.

EXAMPLE

Suppose you are an investigator looking to use the proportional time assumption for your time-to-event study. You want the following for your study:

1 Significance level = 0.05

2	Power	= 0.80
3	Proportional Time Ratio	= 2
5	Loss to follow-up	= 20%
4	One-sided Test	
5	Allocation Ratio	= 1
6	Variance Inflation Factor	= 1

Additionally, you want to estimate the model parameters, k and beta, and you would also like to estimate the proportion of patients who will experience an event during the study based on an accrual time of 12 months and a follow up time of 12 months. You have data from a similar time-to-event study that you have conducted, and you are also interested in obtaining a power curve.

You can input the following into the macro call:

```
%sample_size(Power          = .8,
             alpha          = .05,
             compliance     = .8,
             n_tails        = 1,
             Graph          = 1,
             NumSamples     = 10000,
             k_gam          = "unknown",
             Beta           = "unknown",
             Delta_PT       = 2,
             theta0         = 1,
             r              = 1,
             Evt_prob       = 1,
             r_2            = 0,
             data           = "C:\Users\User\Desktop\example_data.csv",
             d              = "unknown",
             accrual        = 12,
             followup       = 12);
```

The macro will take in prior data that is in a .csv format. Additionally, the data must include the response variable, time, in the first column and the binary variable, event, in the second column. Data from only the control arm is used (even though the prior study data may have had more than two arms), and the proportional time ratio we input defines the effect size comparing the treatment to the control. The macro will calculate model parameters using PROC LIFEREG, and estimate quantiles of the survival function to estimate the overall event rate for the study. Model parameters are calculated as follows:

```
data dummy;
   c = 1;
   output;
run;

data sample;
infile &data
  delimiter=','
  firstobs=2
  missover;
  input time event;
run;

data sample;
set sample dummy;
run;
```

```
ods output ParameterEstimates=Params;
proc lifereg data=sample
model time*event(0)=/dist=gamma;
output out = surv_est quantiles = 0.01 to 0.99 by 0.001
predicted = pred control = c;
run;
data surv_est;
set surv_est;
  sdf = 1-_PROB_;
run;

proc transpose data = params
               out = params1;
id Parameter;
var Estimate;
run;

data params1;
set params1;
  k_gam = 1/Shape**2;
  Beta = abs(Shape)/Scale;
  %if &Beta = "unknown" %then %do;
    call symputx('beta',beta);
  %end;
  %if &k_gam = "unknown" %then %do;
    call symputx('k_gam',k_gam);
  %end;
run;

title 'Parameter Estimation from Prior Study Data';
proc print data = params1;
var k_gam Beta;
run;
```

We read in the prior study data, and provide an indicator variable in the dummy DATA step to be used in the PROC LIFEREG step. This makes sure that we are only predicting the quantiles of the survival function. PROC LIFEREG provides us with estimates of the parameters that we must transform. Additionally, we can take the quantile estimates of the survival function to estimate the event rate. We get the parameter estimates in the following table:

Parameter Estimation from Prior Study Data

Obs	k_gam	Beta
1	3.90147	0.24360

Figure 1. Parameter Estimates

We see that the model parameters have been estimated, and they will now be used in the sample size calculation.

Next the overall event rate must be calculated using Simpson's formula. This will also be done using the prior study data; the macro uses the following formulas to do so:

$$d_i = 1 - \frac{1}{6}\{S_{i(f)} + 4S_{i(f+0.5a)} + s_{i(f+a)}\}$$

$$d = \sum_{i=0}^{1} p_i d_i = \frac{n_0}{N}(d_0 + r d_1)$$

Where *a* and *f* are the accrual and follow-up time, and S_0 is the survival function of the control arm, and S_1 is the survival function of the treatment arm. Also, $N = n_0 + n_1$ is the total number of events. With this estimate of overall event rate, we can adjust our sample size for the study by $N_{total} = \frac{N}{d(r^2)(compliance\ rate)}$.

Next the macro will use the analytic formula to calculate sample size based on the estimated parameters form the data, and the input parameters (Phadnis, Wetmore, & Mayo). This is an iterative procedure that is done within the macro. These values will be the starting values for the estimates of the sample size calculation done via simulation. The program finds an upper and lower bound for the needed sample size and then uses the bisection method to find the needed sample size.

Next the macro will begin the simulations and iterative steps to find the appropriate sample size. The following code simulates the data and is iterated during the macro:

```
data GGamma (keep = SampleID x0 x1 theta0 y arm event);
  theta0=&theta0;
  do SampleID = 1 to &NumSamples;
    do i = 1 to ceil(%sysevalf(&n));
      x0 = rangam(SampleID,&k_gam);
      y = theta0*(x0**(1/&Beta));
      arm = 'Std';
      event = rand('bernoulli',);
      output;
    end;

    do i = ceil(&n*)+1 to ceil(&n*(1+&r));
      x1 = rangam(1000*SampleID,&k_gam);
      y = &Delta_PT*&theta0*(x1**(1/&Beta));
      arm = 'New';
      event = rand('bernoulli',1);
      output;
    end;
  end;
run;

ods exclude all;
ods output ParameterEstimates = ParamTable;

options nonotes;

ods listing;
ods results=off;

proc lifereg data = GGamma;
  by SampleID;
  class  arm ;
  model y*event(0)= arm/dist=gamma maxiter=150;
run;

options notes;
data Table_Mu ;
  set ParamTable (where =(Parameter='arm' and DF=1));
  Est_Mu = Estimate;
run;
```

```
data Table_p (keep=Est_pval signif);
  set ParamTable (where =(Parameter='arm'));
  Est_pval = Probchisq;
  if Est_pval > &alpha/&n_tails then signif = 0;
  else signif = 1;
run;

data PowerCalc;
  set Table_Mu;
  ndf = 2*&n*&k_gam;
  ddf = 2*&n*&r*&k_gam;
  Q = exp(Est_Mu);
  TestStat = ((Q/&Delta_PT)**(&Beta))*&r;
  f_crit = finv(1-(&alpha/&n_tails),ndf,ddf);
  if TestStat > (f_crit*&r)/((&Delta_PT)**(&Beta)) then PctPower = 1;
  else PctPower = 0;
run;

data PowerCalc2;
  set PowerCalc;
  if TestStat = . then delete;
run;

proc means data=PowerCalc2 maxdec=3 N sum mean;
  output out=powercalc3 mean=power1;
  var PctPower;
run;

data powercalc3;
  set powercalc3;
  %if %sysevalf(&check = 0) %then %do;
  call symputx('power1',power1);
run;
  %end;
  %else %do;
  call symputx('power2',power1);
run;
```

The code above simulates data from the generalized gamma ratio distribution. As we can see, it uses the allocation ratio to simulate the correct number of observations for each treatment arm. Next, PROC LIFEREG is used to estimate the parameters and then the following code calculates the test statistic based on the generalized gamma distribution. The macro iterates until if finds the minimum sample size with the requir power.

The macro provides the following output for our example:

K-Gamma	Beta	Desired Power	Power	Overall Event Rate	Compliance	R-Squared	Number of Events	Sample Size Needed
3.90147	0.24360	0.50	0.5056	0.58558	0.8	0	49	105
3.90147	0.24360	0.75	0.7540	0.58558	0.8	0	98	210
3.90147	0.24360	0.80	0.8005	0.58558	0.8	0	112	240
3.90147	0.24360	0.85	0.8532	0.58558	0.8	0	131	280
3.90147	0.24360	0.90	0.9066	0.58558	0.8	0	156	334
3.90147	0.24360	0.95	0.9535	0.58558	0.8	0	195	417

Figure 2. Macro Output

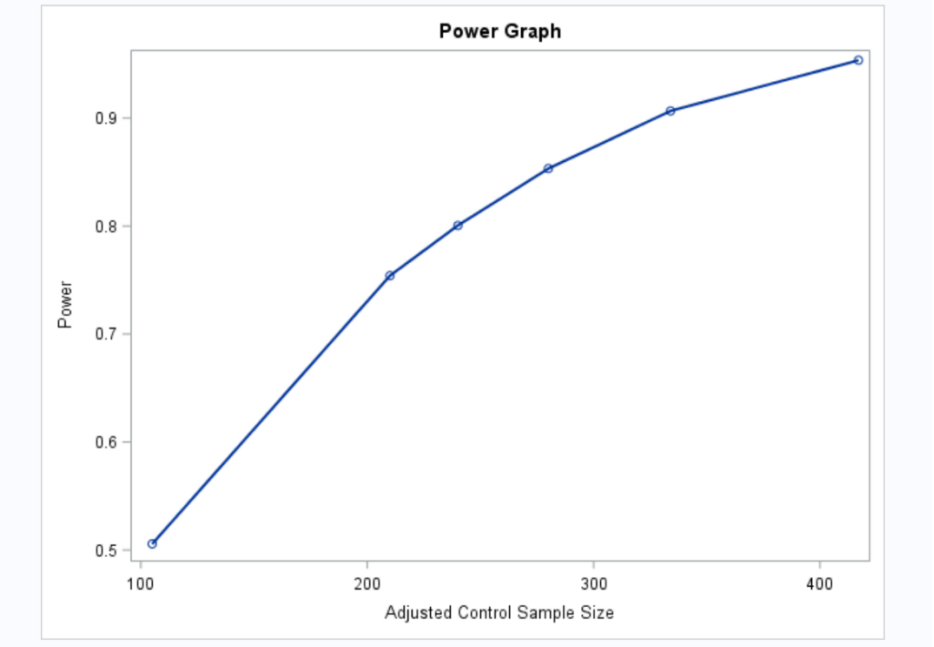

Figure 3. Power Function

The output from the graphing feature provides the power calculated with their associated sample size calculations for the different levels of power, (0.50, 0.75, 0.80, 0.85, 0.90, 0.95), and for the power input by the user.

CONCLUSION

This macro is a flexible sample size calculator for time-to-event studies using the proportional time assumption. By using this macro, investigators will be able to properly power their studies.

REFERENCES

Phadnis MA, Wetmore JB, Mayo MS. A clinical trial design using the concept of proportional time using the generalized gamma ratio distribution. Statistics in Medicine. 2017;36:4121-4140 https://doi.org/10.1002/sim.7421

CONTACT INFORMATION

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

Brian Mosier

University of Kansas Medical Center
bmosier@kumc.edu

Full code is available at:
http://biostat-pts.kumc.edu/velos/jsp/ptsc.jsp