

## Sample size and HR confidence interval estimation through simulation of censored survival data controlling for censoring rate

Giulia Tonini, Menarini Ricerche SPA, Florence (Italy);  
Letizia Nidiaci, Menarini Ricerche SPA, Florence (Italy);  
Simona Scartoni, Menarini Ricerche SPA, Florence (Italy);

### ABSTRACT

Time to event variables are commonly used in clinical trials where survival data are collected. When planning a trial and calculating the sample size, it is important to estimate not only the expected sample size but also the minimum value of the Hazard Ratio for which the trial can be considered successful (i.e. accepting the alternative hypothesis). Calculating the confidence interval for the HR when the difference of treatments is statistically significant, can be performed simulating censored survival data. Simulating censoring can be challenging, especially when the trial does not appropriate to a fixed duration or when high drop-out rate is expected. This is often the case in Oncology trials as patients exit the study prematurely (e.g. in case of AE or disease progression). In this work we present a case study where a dataset is simulated controlling for the resulting censoring rate.

In particular we simulate data from an oncology trial comparing two treatments and provide an example to realize the simulation in SAS. We compare results with varied assumptions regarding drop-out rate and methods to model censoring. Finally, we investigate the effect of the censoring rate on potential bias in estimating the treatment effect using a Cox model.

### INTRODUCTION

Simulation studies for time-to-event data are commonly used in clinical research to mimic trials where the measured outcome is the time until an event occurs. In particular, in oncology trials, mandatory efficacy endpoints are often Overall Survival (OS) or Progression Free Survival (PFS). In those cases the event is death or documented progression of disease respectively. In these kind of studies the event is known, if it happens during the study duration. However, it is not uncommon for patients to exit a study before the event occurs. These incomplete observations in survival data are referred to as censored data. The most common censoring mechanism in clinical research is right censoring, where the event of interest has not occurred by the end of the trial and time to event is known only to be greater than censoring time.

Simulations of these type of trials are useful for example to calculate the minimum hazard ratio for which the null hypothesis is rejected. This can be used to estimate a threshold for the HR when planning a study with the respective sample size.

The distribution of the time to event can be described by either a density function of a parametric distribution or a hazard function. Because of censoring, modeling the time to event is commonly realized through the hazard function. The Cox proportional hazard model is one of the most important statistical models used for censored survival data. It is parametrized in terms of the hazard function and it gives a measure of the correlation between covariates and time to event.

Previous papers have described simulation of survival data from Cox proportional hazards models in a flexible framework [Bender et al.]. However, particular attention should be paid to simulate censoring.

In studies with a fixed duration, censoring can be easily simulated by curtailing the time of observation when the event has not occurred before the fixed time of the study. In several trials (particularly in Oncology trials) that is not the case, since study duration may widely vary from one patient to another and the observation period might have a very large range. Moreover there is often a high drop out rate due to adverse event or other reasons such as patient's choice or detriment to quality of life. For these reasons, simulating censored survival data that reflect a real Oncology trial is not a trivial task. What can be desired is to control for the censoring rate, trying to reproduce a rate that is similar to the observed one in the population of interest.

Realizing such a simulation is an added value for the planning of a study on top of the sample size calculation and it is important especially in the scenario of a development plan when several considerations regarding the possible outcome of a trial need to be done.

## METHODS

The Cox proportional hazards model is given by

$$h(t|x) = h_0(t)\exp(\beta'x)$$

where  $t$  is the time,  $x$  the vector of covariates,  $\beta$  the vector of regression coefficients and  $h_0(t)$  is the baseline hazard function, i.e. the hazard function for  $x=0$ . The effect of the covariates have to be translated from the hazard to the survival times for simulation and analysis. This can be done easily if the baseline hazard is constant, i.e. the survival times are exponentially distributed. This is the reason why the exponential distribution is commonly used in simulations. Another frequently used distribution is the Weibull which is particularly useful when there are discrete covariates, as it is the case for two-arms trials where the covariate of interest is treatment. The Weibull parameters can be chosen such that the hazards are proportional and the true hazard ratio (HR) for the comparison of the two groups can be calculated. In that case, the true regression coefficient for the Cox model can be obtained from  $\log(HR)$ .

The distribution function of the event time is

$$F(t|x) = 1 - \exp(-H_0(t)\exp(\beta'x))$$

Where the cumulative baseline hazard is  $H_0(t) = \int_0^t h_0(t)dt$ .

Let's call  $C$  the censoring time with density function  $\sim g(c|\theta)$ , where  $\theta$  is the censoring parameter. The censoring time can be also modeled with a Weibull distribution. We then let  $y = \min(T, C)$  be the observed follow-up time. If  $T \geq C$  the subject is censored, otherwise the event is observed.

The baseline hazard for a Weibull distribution is

$$h_0(t) = \lambda vt^{\nu-1}$$

Where the cumulative hazard is  $H_0(t) = \lambda t^\nu$ , the density function is  $f_0(t) = \lambda e^{-\lambda t^\nu} \nu t^{\nu-1}$  and the survival function is  $S_0(t) = e^{-\lambda t^\nu}$

Censoring time is simulated from a Weibull distribution, adding a multiplicative factor  $F$  as `Rand("WEIBULL", Seed, (Lambda2) * Linear_Predict*&F.)` [see Code].

In this paper we present the SAS code to realize such simulations. The challenging part of the simulation is ways in which to choose such multiplicative factor  $F$ . Varying the parameter  $F$  gives different scenarios in the simulated data in respect of the resulting censoring rate. We propose an empirical method to choose  $F$  given the target censoring rate.

## CASE STUDY

We consider an oncological study with two arms where we intend to demonstrate superiority of the experimental treatment compared to the best available supportive care. We simulate the dataset with a real effect on survival in order to find the 95%CI of the HR for which the trial can be considered successful, i.e. the null hypothesis is rejected.

We simulate a study with two arms and we keep treatment as the only covariate. We simulate censoring keeping control on the censoring rate. In particular we simulate survival time, censoring time and censoring indicator in the case of rejection of null hypothesis. We then calculate for each simulated dataset, the censoring proportion and we select only the dataset where the censoring proportion is in the desired interval.

In the case study, median survival time for treatment 1 is 10 months, while median survival time for treatment 2 is 14 months, with a HR=0.71. Recruitment time is 24 months and follow up time is 9 months. The sample size is calculated to be 544 patients with 370 events. The observed percentage of censoring is 32%.

We simulate the dataset considering the real values of median survival time from the two treatment groups. In order to reproduce the real case study, we can select the dataset with a censoring rate around between 30% and 35%. We then proceed with a `proc phreg` and we choose only the cases where there is a statistical significance. At that point we calculate the HR with respective confidence interval. The Weibull distribution has a shape parameter equal to 1, which means constant hazard. The scale parameter is  $\lambda e^{trt*\beta}$  where `trt`, the treatment, is the only covariate considered. If  $\beta=1$  the risk is the same in the two groups at time zero. In our case we set  $\beta = \ln(\frac{1}{HR_0})$ . If the treatment is the only covariate, there is no difference in risk at time 0 and `HR_0` is set to 1. We leave  $\beta$  as an input parameter in our macro so that it can be set as one of the initial parameters.

The aim of our work is to understand the sensitivity to the parameter  $F$  in order to choose the best value to simulate the real censoring rate. In Fig 1 we consider the simulation of datasets with a censoring rate interval between 30% and 35%. We run the simulation for different values of the parameter  $F$ . We see that the maximum number of studies with the desired censoring rate is for  $F=2,1$ .

As a sensitivity analysis we repeated the simulation for different target censoring rate. In Fig2 we report the simulation study with a target censoring rate [20%-25%] and we observe an optimal value for  $F$  given by 3,5. In Fig 3, the target censoring rate is [25%-30%] and the optimal value for  $F$  is 2,6.

Under each figure we also report a table with input and output values for the simulation studies. In particular, we report the percentage of studies falling in the target censoring rate and for each of those, the percentage of studies that are actually statistically significant. We then report for each simulation study, the 95% confidence interval of the simulated HR.

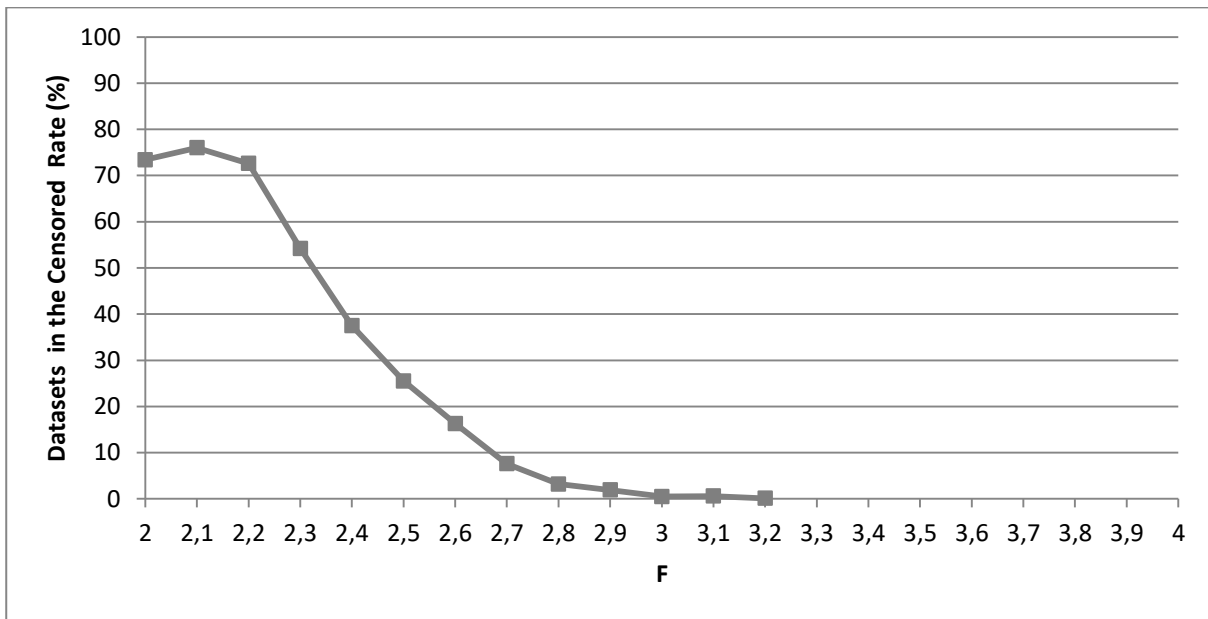
## RESULTS AND DISCUSSION

Our work aims to provide an easy to use simulation code in order to choose the optimal value of the multiplicative factor  $F$  to simulate censoring time from the Weibull distribution given a target censoring rate. In order to have a realistic simulation of the case study,  $F$  can be set to the value giving the maximum number of datasets in the target censoring rate.

The higher limit of the 95%CI of the HR gives us a value for the HR which is still compatible with a trial where the statistical significance is reached for the alternative hypothesis. We observe, from the table of Fig 1, that the 95%CI of the HR changes depending on the number of studies considered and on the percentage of those studies with a  $p$ -value  $<0,05$ . Given that the precision can be increased with a higher number of simulations, this preliminar study gives us an idea of the variability of the HR considering the censoring rate.

We observe that for a lower censoring rate we have a better estimate of the real HR with respect to a study where the censoring rate is high. Performing such simulation studies by varying the censoring rate, can be useful to understand HR expectations from trial results and consequently to understand how close we are to the observation of a real effect.

**Figure 1: Number of datasets (%) in the desired interval for censoring rate [30%-35%] varying the parameter F. (N. simulations =1000; Median 1=10mo; Median 2=14mo; HR=0.71)**

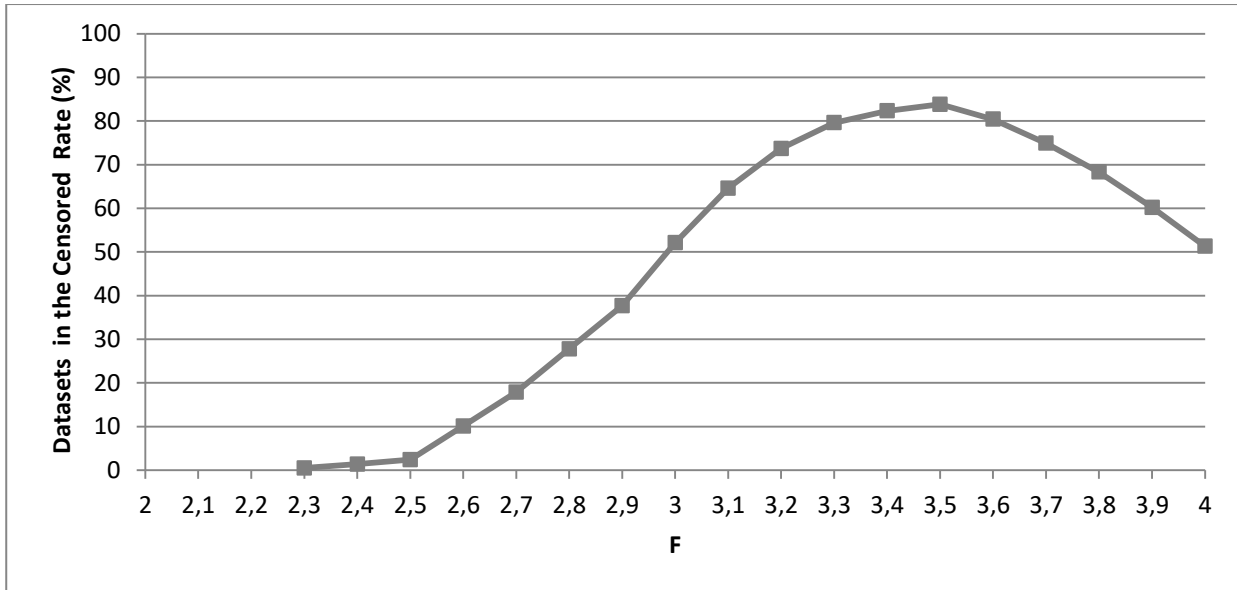


F	Datasets in Censored Rate (%)*	Datasets $p < 0.05$ (%)**	95% LCL for HR	95% UCL for HR
2	73,4	85,7	0,690	0,699
2,1	76	87,9	0,689	0,698
2,2	72,6	87,1	0,694	0,704
2,3	54,2	86,9	0,689	0,700
2,4	37,5	87,7	0,694	0,707
2,5	25,5	85,1	0,696	0,712
2,6	16,3	88,3	0,680	0,702
2,7	7,6	85,5	0,671	0,706
2,8	3,2	87,5	0,679	0,728
2,9	1,9	94,7	0,688	0,731
3	0,5	100	0,641	0,784
3,1	0,6	100	0,640	0,714

\*Percentages are calculated on total number of simulations

\*\* Percentages are calculated on the number of datasets in the censored rate

**Figure 2: Number of datasets (%) in the desired interval for censoring rate [20%-25%] varying the parameter F. ( N. simulations =1000; Median 1=10mo; Median 2=14mo; HR=0.71)**

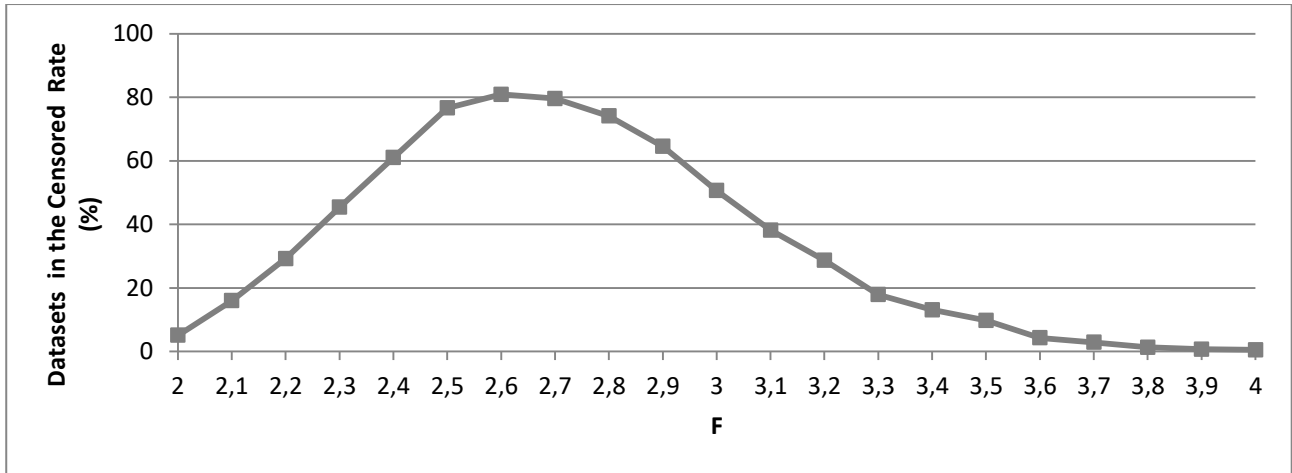


F	Datasets in Censored Rate (%)*	Datasets p<0.05 (%)**	95% LCL for HR	95% UCL for HR
2,3	0,5	80	0,608	0,703
2,4	1,4	100	0,676	0,776
2,5	2,4	91,7	0,655	0,715
2,6	10,1	88,1	0,687	0,710
2,7	17,9	91,6	0,692	0,710
2,8	27,8	88,5	0,694	0,710
2,9	37,7	90,7	0,689	0,703
3	52,1	88,3	0,698	0,710
3,1	64,6	90,2	0,701	0,711
3,2	73,7	89,7	0,702	0,711
3,3	79,6	89,8	0,702	0,711
3,4	82,3	90,4	0,698	0,706
3,5	83,8	90,1	0,702	0,710
3,6	80,4	91,5	0,699	0,708
3,7	74,9	90,6	0,696	0,705
3,8	68,3	90,3	0,698	0,708
3,9	60,2	90,5	0,699	0,709
4	51,3	90,2	0,696	0,707

\*Percentages are calculated on total number of simulations

\*\* Percentages are calculated on the number of datasets in the censored rate

**Figure 3: Number of datasets (%) in the desired interval for censoring rate [25%-30%] varying the parameter F. ( N. simulations =1000; Median 1=10mo; Median 2=14mo; HR=0.71)**



F	Datasets in Censored Rate (%)*	Datasets p<0.05 (%)**	95% LCL for HR	95% UCL for HR
2	5,1	92,1	0,673	0,714
2,1	16	87,5	0,684	0,705
2,2	29,2	88,7	0,688	0,702
2,3	45,4	85,7	0,695	0,707
2,4	61	87,7	0,696	0,705
2,5	76,6	88,4	0,694	0,702
2,6	80,9	89,7	0,698	0,707
2,7	79,6	88,9	0,700	0,709
2,8	74,1	88,2	0,694	0,703
2,9	64,5	88,4	0,698	0,708
3	50,6	86,6	0,688	0,699
3,1	38,2	89,5	0,692	0,705
3,2	28,7	89,2	0,694	0,708
3,3	17,9	90,5	0,687	0,706
3,4	13,1	87,8	0,685	0,705
3,5	9,8	90,8	0,703	0,728
3,6	4,3	90,7	0,682	0,726
3,7	2,9	93,1	0,645	0,697
3,8	1,3	84,6	0,665	0,748
3,9	0,7	100	0,606	0,735
4	0,5	80	0,682	0,767

\*Percentages are calculated on total number of simulations

\*\* Percentages are calculated on the number of datasets in the censored rate

## CODE

```
%macro Simulation(
median1 = ,
median2 = ,
F=,
Beta=,
N_sim=,
N_pz=,
cens_L=,
cens_U=,
out= );
%let HR=%SYSEVALF(&median1./&median2.);
data pop;
  Seed= 1 ;
  censored=.;
  Beta1=&Beta.;
  Lambda2=1/(log(2)/&median2.);
  Lambda1=1/(log(2)/&median1.);
  Do Sim = 1 To &N_sim. ;
    Do Trt = 1 To 2 ;
      Do _n_ = 1 To &N_pz. ;
        ID+1;
        if TRT=1 then do;
          Linear_Predict = Exp(-Beta1 * Trt ) ;
          Time = Rand( "WEIBULL", Seed, (Lambda1) * Linear_Predict) ;
          Time_censor = Rand( "WEIBULL", Seed, (Lambda1) * Linear_Predict*&F.)
;
          Time_Days = Ceil( Time * 30 ) ;
          end;

          if TRT=2 then do;
            Linear_Predict = Exp(-Beta1 * Trt ) ;
            Time = Rand( "WEIBULL", Seed, (Lambda2) * Linear_Predict) ;
            Time_censor = Rand( "WEIBULL", Seed, (Lambda2) * Linear_Predict*&F.)
;
            Time_Days = Ceil( Time * 30 ) ;
            end;
            if time_censor <= time then do;
              censored=1;
              time=time_censor;
            end;
            if time_censor > time then
              censored=0;
            Output ;
          End ;
        End ;
      End ;
    End;
  drop Linear_Predict;
Run ;

/*Tot of censoring x study*/
proc sql;
create table cens_N as
select count(censored) as cens_study, sim
from pop
where censored=1
group by sim;
run;
quit;
```

```

data pop2;
merge pop cens_N;
by Sim;
if cens_study=. then cens_study=0;
format cens_perc 8.2;
cens_perc=(cens_study/(&N_pz.*2));
if &cens_L.<=cens_perc<=&cens_U. then SIMFL=1; else SIMFL=0;
run;

/*Select studies in the target censoring rate */
data pop3;
set pop2 (where=(SIMFL=1));
run;
proc sort data=pop3;by Sim;run;

data pop3_N;
set pop3;
by sim;
if first.sim;
run;
proc sql;
select count(sim) into :FL_tot
from pop3_N;
run;
quit;

/*Dataset summary by simulation*/
data summary_bysim_&out.;
set pop2;
by sim;
if first.sim;
cens_L=&cens_L.;
cens_U=&cens_U.;
F=&F.;
L2=&median2.;
L1=&median1.;
T=&FL_tot.;
P=(T/&N_sim.)*100;
Label sim='N sim' cens_study='N Censored' cens_perc='% Censored' L2='Median 2'
L1='Median 1' T='Tot Studies in the Cens Rate'
cens_L='Lower limit % Cens' cens_U='Upper limit % Cens' P='% Studies in the Cens
Rate';
keep sim cens_study cens_perc F L2 L1 T cens_L cens_U P ;
run;

/*Model*/
Ods Output ParameterEstimates=out_PHreg;
Proc PHReg Data = pop3 ;
Model Time*Censored( 1 )=Trt / RiskLimits = Wald;
By Sim ;
Run;

/*Select the HR of stat significant sim */
data out_PHreg_1;
set out_PHreg;
if ProbChiSq<=0.05;
run;

proc sql;
select count(sim) into :S_tot
from out_PHreg_1;
run;
quit;

```



```

/* 95% CI of HR*/
ods output summary=CI_HR;
Proc Means Data = out_PHreg_1 alpha=0.05 CLM;
  Class Parameter ;
  Var HazardRatio;
Run ;

/*Dataset summary overall*/
data last;
set summary_bysim_&out. end=eof;
if eof then output;
drop cens_study cens_perc;
run;

data sim.summary_&out.;
set last;
set CI_HR;
PS=(NObs/&FL_tot.)*100;
HR=&hr.;
label NObs='Studies p<0.05'
PS='% Studies p<0.05';
drop Parameter;
run;

proc datasets NOLIST; delete Cens_n ci_hr last out_phger out_phreg_1 pop pop2
pop3 pop3_n out_phreg; run; quit;
%mend;

%macro combine (max);
data sim.summary_tot;
  set
  %do i = 1 %to &max.;
    sim.summary_&i.
  %end;
;
run;
%mend;

/*****/

/*Target cens rate: (0.30;0.35) and F=2*/
%Simulation(
median1 = 10,
median2 = 14,
F=2,
Beta=log(1),
N_sim=1000,
N_pz=250,
cens_L=0.30,
cens_U=0.35,

OUT=1 );

```

## REFERENCES

Bender A, Altman DG, Royston P, Holder RL. 2006. "Generating serviva times to simulate Cox proportional hazard models" *Statistics in medicine*, 25(24):4279-4292

## CONTACT INFORMATION

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

Giulia Tonini

Menarini Ricerche

Via Rosolino Pilo 10

50131 Firenze, Italy

Email: [gtonini@menarini-ricerche.it](mailto:gtonini@menarini-ricerche.it)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.