



## Simulations-Based Least Required Sample Size and Power in Clinical Trials with Time-to-Event endpoint and Variable Hazard

Mohamed Mubasher<sup>1\*</sup>, Brian M. Rivers<sup>2</sup>, Liang Shan<sup>3</sup>, Fengxia Yan<sup>4</sup>, Fan Wu<sup>5</sup>, Muhammed Idris<sup>6</sup>, Alexander Quarshie<sup>7</sup>, Robert M. Mayberry<sup>8</sup>, Elizabeth Ofili<sup>9</sup>, Tabia Henry Akintobi<sup>10</sup>, Sejong Bae<sup>11</sup>.

<sup>1</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

<sup>2</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine.

<sup>3</sup>Department of Medicine, Division of Preventive Medicine O'Neal Comprehensive Cancer Center UAB | University of Alabama at Birmingham.

<sup>4</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

<sup>5</sup>Computer Science Department, Tuskegee University Tuskegee, AL.

<sup>6</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

<sup>7</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

<sup>8</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

<sup>9</sup>Department of Medicine (cardiology) Morehouse School of Medicine, Atlanta GA.

<sup>10</sup>Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

<sup>11</sup>Department of Medicine, Division of Preventive Medicine O'Neal Comprehensive Cancer Center UAB | University of Alabama at Birmingham.

### Article Info

**Received:** September 04, 2024

**Accepted:** September 07, 2024

**Published:** September 20, 2024

**\*Corresponding author:** Mohamed Mubasher, Community Health and Preventive Medicine Department Morehouse School of Medicine, Atlanta GA.

**Citation:** Mubasher M, Brian M. Rivers, Liang Shan, Yan F, and Wu F, et al. (2024) "Simulations-Based Least Required Sample Size and Power in Clinical Trials with Time-to-Event endpoint and Variable Hazard.". International Journal of Epidemiology and Public Health Research, 5(2); DOI: 10.61148/2836-2810/IJEPHR/083

**Copyright:** © 2024 Mohamed Mubasher. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract:

Two of the pivotal design parameters for planning clinical trials with time-to-event outcome(s) are sample size and power. Attention needs to be placed on the hazard function (which characterizes the rate at which events occur and can be constant, decreasing, and/or increasing in time). This work employs simulation(s) of real scenarios of randomized studies to generate time-to-event variables with specific hazard characterization, obeying the Weibull function which accommodates variable hazard situations. Our aim is to determine the least required sample size and power values, based on simulating two independent samples of Weibull distributed responses, differing by various postulated hazard patterns (constant, decreasing, or increasing in time), different scale parameter values, and follow-up periods.

**Keywords:** time-to-event, simulation, hazard, Weibull, statistical power

### 1.0 Introduction and background:

Sample size and power are vital parameters in planning and designing experimental research in general and particularly for randomized clinical trials.<sup>1</sup> Among the challenges, is the need to establish a meaningful and relevant difference (or effect size) to be detected, along with a quantifiable, expected variability that characterizes the response (outcome) variable. In planning studies with time-to-event variables, e.g., time to failure and/or disease progression as primary outcome(s), special attention must be placed on characterizing the rate at which events occur in relation to some measure (e.g. time, space), which is also called hazard<sup>2</sup>. Hazard can be constant, decreasing, and/or increasing over time. To that end, the Weibull function<sup>2</sup> can be employed as a very useful and versatile statistical distribution to generate time-to-event variables with specific hazard characterization. It is equally important to realize that statistical power (the probability of detecting a significant shift in time-to-event (also called generically, survival) for time-to-event outcomes is also a direct function of the expected number of events, which in turn, is determined by the hazard function.

## Time-to-event (survival) outcomes in power and sample size calculations

Schoenfeld<sup>3</sup> derived the asymptotic distribution under alternative hypotheses for a class of statistics, including the log-rank and the modified Wilcoxon statistic, for testing the equality of two survival distributions in the presence of arbitrary right censoring and proportional hazard function. Lachin<sup>4</sup> showed the general formula for sample size determination when testing equality of hazard functions for experimental and control groups, under the assumption that time to survival is exponentially distributed. A few years later, Lachin and Foulkes<sup>5</sup> generalized the method to encompass non-uniform accrual, loss to follow-up, noncompliance, and stratification. Schoenfeld<sup>6</sup> extended his previous work (Schoenfeld<sup>3</sup>) to derive the asymptotic power of a class of statistics to test the equality of two survival distributions, by incorporating subject-specific concomitant covariates within the proportional-hazards model framework. Palta and Amini<sup>7</sup> studied the effect of stratification and covariates on sample size determination. A more theoretical work was developed by Lakatos<sup>8</sup>, which provided a proportional hazard-independent method, using binomial proportions and the Tarone-Ware class of statistics, to estimate sample sizes for the comparison of survival distributions based on the log rank statistic, in the presence of strata-specific noncompliance and lag time rates. To generalize Schoenfeld's derivation<sup>6</sup> and the work of Makuch and Simon<sup>9</sup>, Ahnn and Anderson<sup>10</sup> presented an algorithm for comparing two survival distributions, based on a sample size formula for testing monotone dose-response using Tarone's trend test and the stratified log rank test. Others like Machin et al<sup>11</sup> also provided computational algorithms for survival distribution comparisons, based on Lachin's work<sup>4</sup> and/or other normal approximation methodology. Unlike Lakatos<sup>8</sup>, Hsieh and Lavori<sup>12</sup> used a proportional hazards regression model with a non-binary covariate to derive a formula to produce the required number of deaths/events, which does not require pre-specification of the time-to-event distribution. In addition, their procedure showed that censored observations do not contribute to the power of the test in the proportional hazards model and allow for adjustment of sample size in the presence of additional covariates. Although Castleoe and O'Brien<sup>13</sup> discussed analytical components and methods for approaching sample size and power for various simple linear models, they were mainly focusing on normally distributed outcomes and did not include time-to-event responses. Lakatos<sup>14</sup> presented an adaptive sequential methodology for designing complex group sequential survival trials, when the survival curves are compared using the log rank statistic. Shen and Cai<sup>15</sup> proposed a flexible adaptive design that continuously updates sample size in clinical trials for censored survival data with staggered study entry and random loss to follow-up. This test statistic is based on the weight (by prior observed data) of the linear log rank. Their approach allows for early termination and acceptance of the null hypothesis, when no difference is detected between treatment groups, however, the null hypothesis may only be rejected at the last step, when the weight function is used up.

Song et al<sup>16</sup> developed robust, covariate-adjusted, log-rank statistics applied to recurrent events data with arbitrary numbers of events under independent censoring and the corresponding sample

size formula. Zhang and Quan<sup>17</sup> derived formulas necessary to calculate the asymptotic power of the log-rank test under this non-proportional hazards' alternative for two data analysis strategies. Simulation studies indicated that the formulas provide accurate power for a variety of trial settings. In Lachin<sup>18</sup>, general expressions are described for the evaluation of sample size and power for the K group Mantel-log-rank test or the Cox proportional hazards (PH) model score test, and for a stratified-adjusted K group test and for the assessment of a group by stratum interaction. Chen et al<sup>19</sup> developed a sample size determination method for the shared frailty model to investigate the treatment effect on multivariate event times.

### Modeling time-to-event outcomes, according to the Weibull distribution :

Although the Weibull distribution encompasses more general cases of survival times, only a few have considered that in power analysis. Heo et al<sup>20</sup> are among the first to consider power and sample size under Weibull distribution assumption, and their work has shown that power and sample sizes are heavily dependent on the shape parameter of the Weibull distribution. Jeske et al<sup>21</sup> developed a procedure for sample size calculations that was constrained by specified Type-1 and Type-2 error rates for comparing Weibull-distributed survivor functions of multiple treatment groups versus a control group that also allowed for Type-1 censoring. Wu<sup>22</sup> proposed two parametric tests for designing randomized, two-arm, Phase III survival trials under the Weibull model and further investigated the effects of misspecification of Weibull parameters on sample size estimates. Lim et al<sup>23</sup> reported a comparative account of methods for sample size estimation under the Weibull distribution, from a reliability standpoint for applications in the engineering field. They compared the accuracy of reliability measures calculated by conventional statistical and Bayesian methods, and accordingly, devised a procedure comprised of six steps comparing the accuracy of prior information, but also considering the failure model, sample size, and censoring ratio. Yang et al<sup>24</sup> compared an engineering field application, relying on the computing sample size estimated according to the two-parameter Weibull distribution and compared their performance to that based on a normal distribution.

To the best of our knowledge, no published literature nor available software tools (e.g., SAS, PASS, nQuery) for power analysis in survival analysis, thus far, has explicitly and directly focused on sample size and power, as being influenced by the shape of the hazard function and the scale parameter in the Weibull function. In this work, we determined the least required sample size and power (LRSSP), based on simulating two independent samples of Weibull-distributed responses, differing by various postulated hazard ratios (HR) and characterized by either constant, decreasing, or increasing in time hazard patterns; different scale parameter values; and follow-up periods.

### 2.0 Methods:

Simulation is used to estimate the LRSSP, in relation to varying outcome responses and corresponding hazard patterns in clinical trials. Specifically, time-to-event random variables, obeying

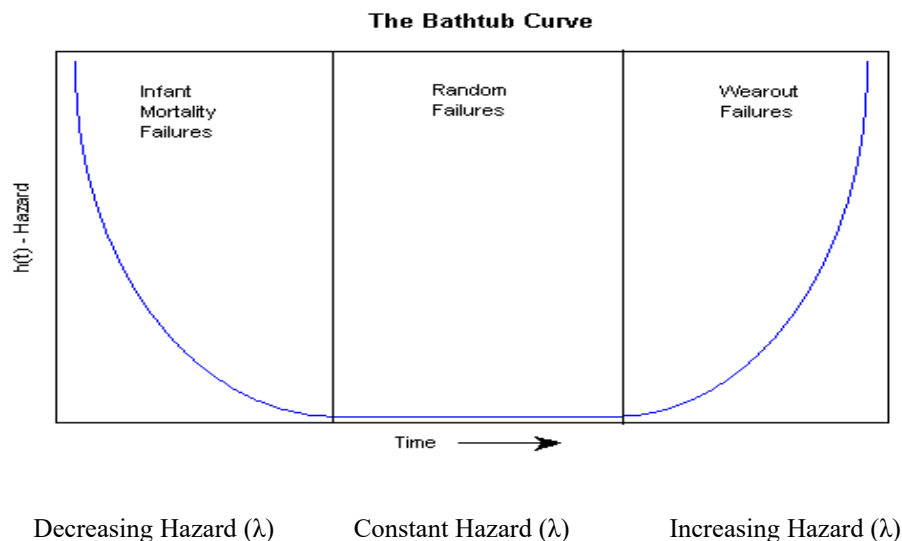
Weibull distribution, were simulated under different hazard patterns and scale values, sample sizes, deltas (shift in distribution under the alternative hypothesis, as a result of specific value of HR), and follow-up parameters. The probability density function of a Weibull random variable is given by:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where  $k > 0$  is the shape parameter, and  $\lambda > 0$  is the scale parameter.

The Weibull distribution is known to mimic time-to-event phenomena. It is the general case of the exponential distribution ( $k = 1$ ). The failure rate for the Weibull is proportional to power of time (power is the shape parameter plus one). Shape parameter ( $k$ ) of values  $< 1$  represent decreasing failure rate over time (e.g., when there is significant presence of "infant mortality", or defective items early on during monitoring/follow-up),  $k = 1$  corresponds to constant hazard over time (i.e., presence of random external factors impacting failure), and  $k > 1$  indicates increasing hazard in time (e.g., aging/wearing off effect), creating "The Bathtub Curve" (Figure 1).

Figure 1<sup>1</sup>



## 2.1 Simulation process and parameters

### Study Design

Two randomized study treatment groups were chosen for this work (reference and experimental), differing by a shift in the time-to-event distribution. In this work, the *outcome response* variable is defined as time-to-event/failure, denoted by  $x_{kij}$  for the  $k^{\text{th}}$  subject in the  $i^{\text{th}}$  study group at time  $t_j$ , which was assumed to follow a (two parameter) Weibull distribution. An example of such a scenario is a two-arm, randomized cancer clinical trial (chemotherapy alone vs. chemotherapy and immunotherapy), comparing time to relapse during a three year follow-up period.

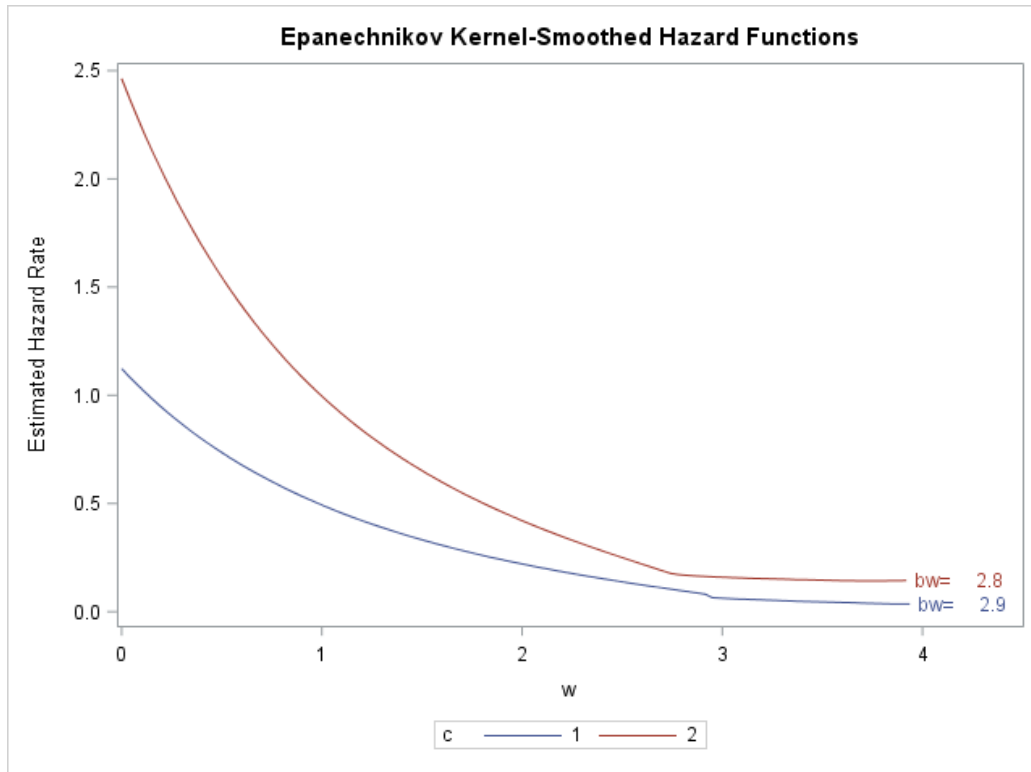
### Simulation parameters

a) We selected Weibull sample shape parameter values ( $k$ ) of 0.8, 1, and 2.0, to respectively, represent decreasing, constant, and increasing hazard functions in time (Figures 2, 4, and 5), respectively. Scale ( $\lambda$ ) parameters of 2.5 and 4.5 were chosen to closely mimic years as a measuring unit and simultaneously varying the dispersion parameters for the distribution. The distribution under the null of equality of the two groups' survival distributions ( $H_0: S(t)_1 = S(t)_2$ ) is assumed to be asymptotically equivalent for different values of  $k$  and  $\lambda$ .

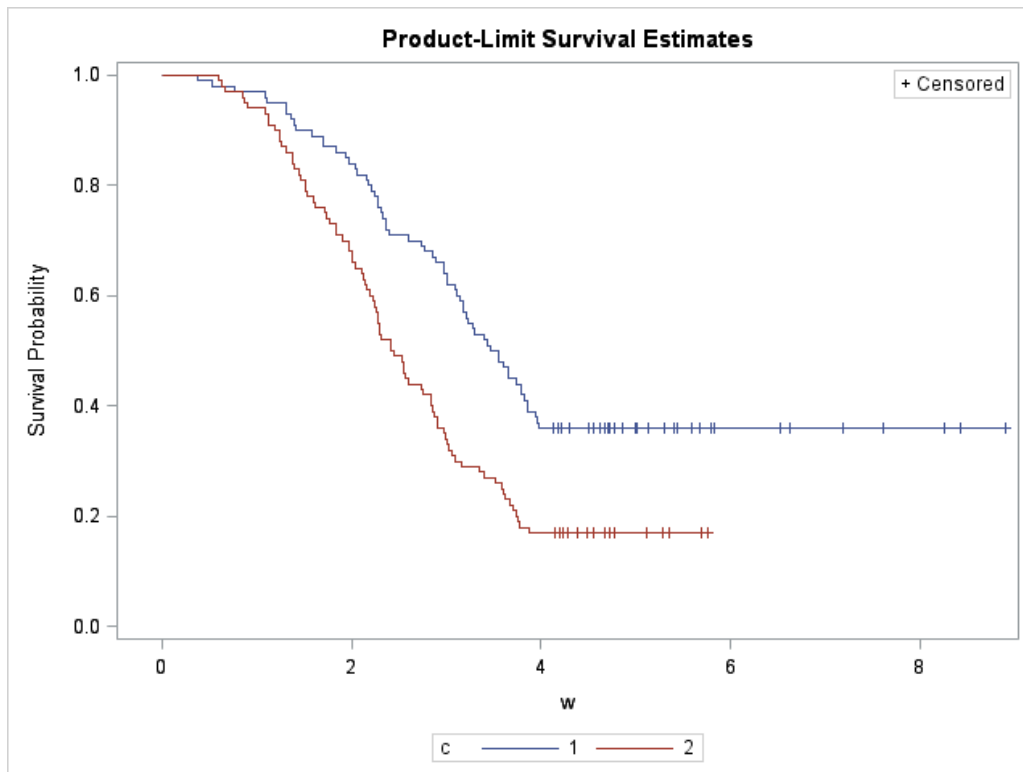
b) Follow-up periods ( $M_i$ ) were set at 3, 4, and 5 years (for example).

c) Sample size for each group was investigated for the values of 30, 40, 50, 75, 100, 150, 200, 250, and 300.

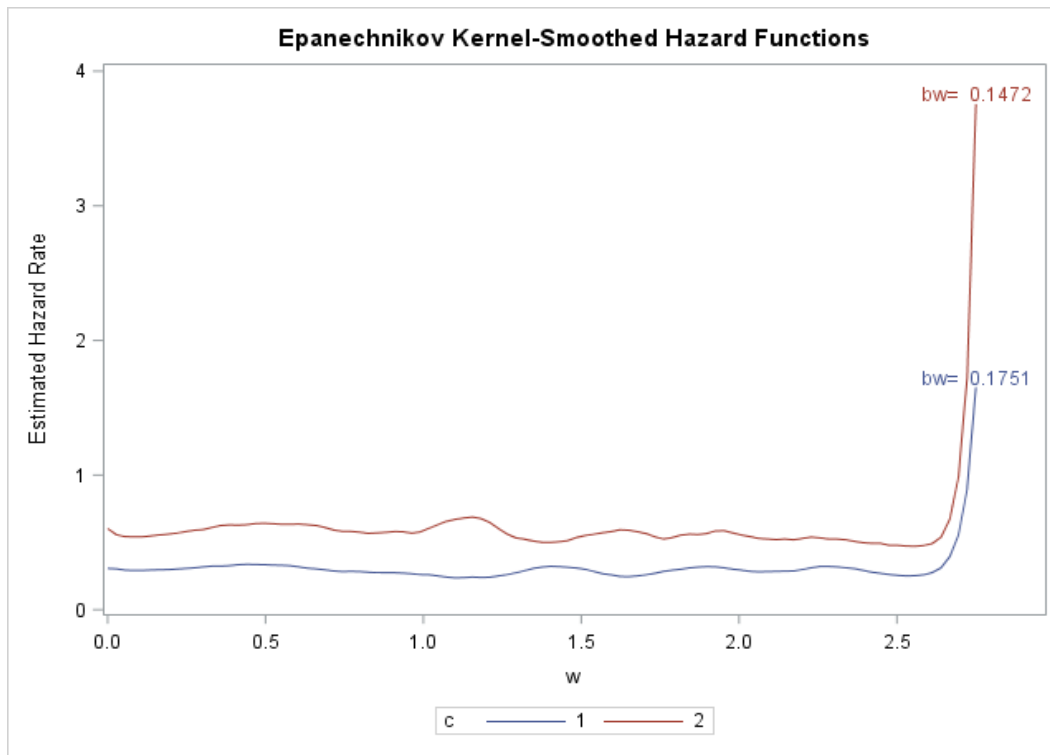
d) To produce alternative hypotheses ( $H_a$ ), we selected HR values of 0.80, 0.70, 0.65, 0.60 and 0.50, which were devised employing the simplest form of Cox proportional hazard model ( $h(\mathbf{Z}; x) = h_0(x) \exp(\mathbf{Z}\beta)$ ). In this instance, the only covariate,  $\mathbf{Z}$ , is study (treatment) group. Note, that these HR values correspond to deltas (reduction in hazard) equivalent to 20%, 30%, 35%, 40%, and 50%, respectively.



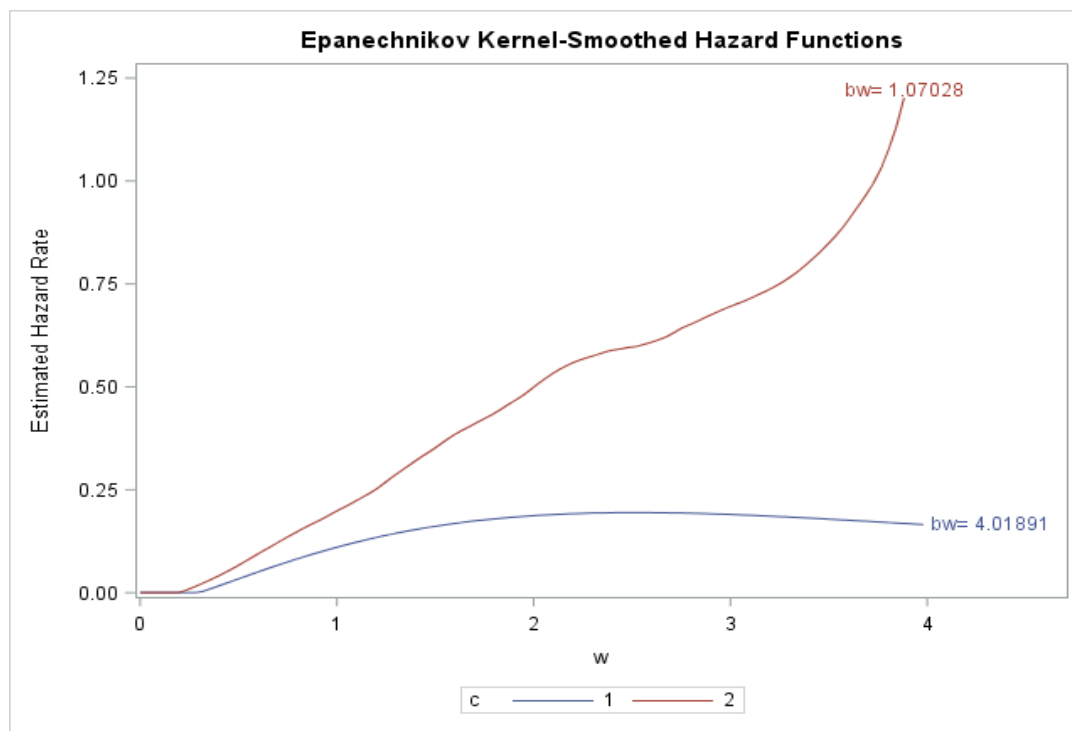
**Figure 2:** Typical hazard functions of Simulated Sample size 100 per group, Decreasing Hazard (shape=0.8), HR= 0.5, Follow-up=4



**Figure 3:** Typical Survival distributions with Simulated Sample size 100 per group, shape=2.0, HR= 0.5, Follow-up=4



**Figure 4:** Typical hazard functions of Simulated Sample size 200 per group, Constant Hazard (shape=1.0), HR= 0.5, Follow-up=5



**Figure 5:** Typical hazard functions of Simulated Sample size 100 per group, Increasing Hazard (shape=2.0), HR= 0.5, Follow-up=4 Years

## 2.2 Computation of statistical power

Each study participant has an associated time-to- event,  $x_{lij}$ , which at follow-up cutoff value ( $M_i$ ) can either be a censored observation

(if the failure time,  $x_{lij}$ , is greater than  $M_i$ ) or otherwise, an event time (if  $x_{lij}$  is less than or equal to  $M_i$ ). The log rank test was then applied to test equality of survival distributions. This entire process

was repeated 5,000 times for each value of the simulation parameters, for a total of 1,500,000 replicas.

At each of these simulation iterations, power, the probability of Figure 2 illustrates the shift in survival function for the experimental group in a typical simulation iteration, and Table 1 exemplifies distribution cut points with sample size 300 for each

rejecting  $H_0$  under  $H_1$ , was calculated under the various HR values [c.f. section d) above] by simply computing proportion of rejections using a fixed 5% significance level ( $\alpha = 0.05$ ).

group, shape parameters equivalent to 1.0, scale = 2.5, and follow-up period of 5 years.

Study Group	N	Mean	Std Dev	Min	Percentiles					Max
					10th	25th	50th	75th	90th	
1	1500000	2.50	2.50	0.00	0.26	0.72	1.73	3.46	5.76	35.20
2	1500000	2.00	2.00	0.00	0.21	0.58	1.39	2.78	4.61	32.08

**Table 1:** Distributional Cut points of Simulated Time-to-event by Study Group (Sample size 300 on each group, shape parameters=1.0, scale =2.5 and follow-up period=5 years)

### 2.3 Least Required Sample Size and Power (LRSSP)

Power values and sample size are interlaced. Generally, power is a function of sample size, HR, and variability, as well as follow-up.

1. Given  $H_1$ , the algorithm for the simulation process goes as follows: Specify, using a macro facility in SAS, the values of the simulation parameters (sample size, Weibull shape and scale, delta for the Cox model, and follow-up time) and number of samples to be created.
2. Simulate Weibull random variables according to (1) under the alternative hypothesis.
3. Compute log rank statistic.
4. Test for significance using  $\alpha = 0.05$  and corresponding chi-square cutoff value.
5. Compute proportion of the number of rejections.
6. Therefore, LRSSP is iteratively determined as the solution of a power value of at least 0.70 and the corresponding values of HR, shape and scale parameters, and follow-up time.

### 2.4 Computational tools

Procedure Interactive Matrix Language (IML) in SAS (v.9.4; SAS Institute, Inc., Cary, NC) was the tool of choice to execute these simulations, along with Lifetest and frequency procedures.

Simulations' results were confirmed using R version 4.3.0.

### 3.0 Results

Tables 2, 3, and 4 of LRSSP correspond to decreasing, constant, and increasing hazard in time, respectively. It is evident that power is a monotonic, non-decreasing function of increasing sample size (as it should!), reduction in hazard, and increasing follow-up time, as well as a decreasing scale. LRSSP can be read directly from these tables, given: a) anticipated overall shape of hazard (decreasing, constant, or increasing); b) desired power; c) anticipated reduction in hazard for the experimental group, relative to the reference group; and d) the specifically anticipated value for the scale parameter (which impacts variability). For example, in a 3 year follow-up study where time-to-event is expected to have a decreasing hazard in time with scale value of 2.5, 90% desired power to detect a 35% relative reduction in hazard for the tested group, the LRSSP per group is 150 subjects. However, with increasing hazard (Weibull shape parameter = 2) and given the same parameters, sample size of 150 per group will afford 92% power. Similarly, one can determine that a sample size per group of 75 subjects will suffice to produce 81% power, under constant hazard function, scale parameter value of 2.5, HR of 0.60, and follow-up duration of 5 years.

	Hazard ratio (Reduction in hazard for the experimental group)									
	0.80 (20%)		0.70 (30%)		0.65 (35%)		0.60 (40%)		0.50 (50%)	
	2.5	4.5	2.5	4.5	2.5	4.5	2.5	4.5	2.5	4.5
Least Required Sample Size (per group)										
30	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.669 0.710 0.724	<0.70
40	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.784 0.815 0.833	0.700 0.748 0.778

50	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.869 0.894 0.906	0.789 0.835 0.860
75	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.776 0.812 0.834	0.674 0.730 0.762	0.968 0.977 0.980	0.923 0.947 0.963
100	<0.70	<0.70	<0.70	<0.70	0.750 0.785 0.811	0.648 0.698 0.737	0.882 0.906 0.919	0.805 0.847 0.874	>0.99	0.973 0.987 0.992
150	<0.70	<0.70	0.758 0.794 0.813	0.646 0.712 0.750	0.897 0.917 0.932	0.812 0.8630. 890	0.967 0.979 0.986	0.927 0.952 0.964	>0.99	>0.99
200	<0.70	<0.70	0.867 0.893 0.905	0.764 0.827 0.854	0.961 0.972 0.978	0.909 0.938 0.956	>0.99	0.974 0.987 0.992	>0.99	>0.99
250	<0.70	<0.70	0.932 0.951 0.959	0.857 0.902 0.923	0.989 0.992 0.995	0.959 0.975 0.985	>0.99	>0.99	>0.99	>0.99
300	0.645 0.689 0.712	<0.70	0.966 0.975 0.982	0.912 0.943 0.958	>0.99	0.983 0.991 0.994	>0.99	>0.98	>0.99	>0.99

**Table 2:** Least Required Sample Size and Power values as result of various postulated reductions in hazard when failure time distribution is characterized by decreasing Hazard in Time (Weibull Shape parameter=0.8)

	Hazard ratio (Reduction in hazard for the experimental group)									
	0.80 (20%)		0.70 (30%)		0.65 (35%)		0.60 (40%)		0.50 (50%)	
	2.5	4.5	2.5	4.5	2.5	4.5	2.5	4.5	2.5	4.5
Least Required Sample Size (per group)										
30	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.678 0.714 0.733	<0.70
40	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.789 0.821 0.839	0.681 0.743 0.782
50	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.871 0.900 0.913	0.767 0.832 0.861
75	<0.70	<0.70	<0.70	<0.70	0.629 0.678 0.705	<0.70	0.781 0.823 0.846	0.654 0.724 0.769	0.969 0.977 0.981	0.912 0.947 0.965
100	<0.70	<0.70	<0.70	<0.70	0.755 0.799 0.823	0.626 0.699 0.740	0.886 0.912 0.927	0.789 0.842 0.877	>0.99	0.968 0.986 0.992
150	<0.70	<0.70	0.766 0.804 0.829	0.628 0.703 0.750	0.903 0.927 0.941	0.792 0.858 0.892	0.969 0.983 0.988	0.918 0.950 0.967	>0.99	>0.99
200	<0.70	<0.70	0.871 0.900 0.918	0.749 0.818 0.859	0.964 0.975 0.981	0.893 0.936 0.958	>0.99	0.966 0.987 0.993	>0.99	0.99
250	<0.70	<0.70	0.934 0.954	0.841 0.896	>0.98	0.952 0.975	>0.99	>0.99	>0.99	>0.99

			<b>0.967</b>	<b>0.926</b>		<b>0.986</b>				
<b>300</b>	<b>0.649</b> <b>0.702</b> <b>0.728</b>	<b>&lt;0.70</b>	<b>0.968</b> <b>0.981</b> <b>0.984</b>	<b>0.900</b> <b>0.940</b> <b>0.960</b>	<b>&gt;0.99</b>	<b>&gt;0.98</b>	<b>&gt;0.99</b>	<b>&gt;0.98</b>	<b>&gt;0.99</b>	<b>&gt;0.99</b>

**Table 3:** Least Required Sample Size and Power values as result of various postulated reductions in hazard when failure time distribution is characterized by Constant Hazard in Time (Weibull Shape parameter=1.0)

Scale	Hazard ratio (Reduction in hazard for the experimental group)									
	0.80 (20%)		0.70 (30%)		0.65 (35%)		0.60 (40%)		0.50 (50%)	
	2.5	4.5	2.5	4.5	2.5	4.5	2.5	4.5	2.5	4.5
Least Required Sample Size (per group)										
30	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.708 0.745 0.748	<0.70
40	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.816 0.848 0.851	0.562 0.723 0.793
50	<0.70	<0.70	<0.70	<0.70	<0.70	<0.70	0.653 0.701 0.710	<0.70	0.893 0.918 0.921	0.651 0.808 0.876
75	<0.70	<0.70	<0.70	<0.70	0.6650. 0.716 0.729	<0.70	0.716 0.773 0.810	0.542 0.698 0.786	0.943 0.968 0.977	0.819 0.936 0.972
100	<0.70	<0.70	0.624 0.689 0.701	<0.70	0.782 0.839 0.844	0.500 0.670 0.757	0.906 0.935 0.943	0.654 0.821 0.889	>0.99	0.918 0.980 0.993
150	<0.70	<0.70	0.792 0.847 08.57	0.500 0.672 0.769	0.917 0.947 0.950	0.676 0.839 0.906	0.979 0.990 0.991	0.826 0.938 0.973	>0.99	0.826 0.938 0.973
200	<0.70	<0.70	0.893 0.931 0.937	0.627 0.789 0.874	0.971 0.987 0.989	0.793 0.923 0.966	>0.99	0.916 0.980 0.994	>0.99	>0.99
250	0.616 0.682 0.700	<0.70	0.952 0.973 0.977	0.727 0.879 0.938	>0.99	0.883 0.966 0.991	>0.99	0.959 0.996 0.999	>0.99	>0.99
300	0.689 0.746 0.768	<0.70	0.976 0.986 0.990	0.801 0.924 0.969	>0.98	0.931 0.986 0.996	>0.99	>0.98	>0.99	>0.99

**Table 4:** Least Required Sample Size and Power values as result of various postulated Reductions in hazard when failure time distribution is characterized by Increasing Hazard in Time (Weibull Shape parameter=2.0)

#### 4.0 Discussion

##### Conclusion

This simulation work demonstrates that sample size and power can be precisely determined, using the features of the Weibull distribution which allow for control of the shape of the hazard function and scale parameter. These results incorporate additional structure specifications of time-to-event distributions under the alternative hypothesis that potentially enhance sample size computation accuracy.

##### Rationale for selection of simulation parameters

In time-to-event studies, the risk/hazard cannot always be absolutely determined and does not necessarily stay systematic throughout follow-up/observation period (i.e., fluctuates in direction and magnitude). Nonetheless, it is imperative, in designing such studies, to have a realistic and informed expectation of this parameter from pilot and/or preliminary data, existing literature, or simply from intuitive knowledge of the

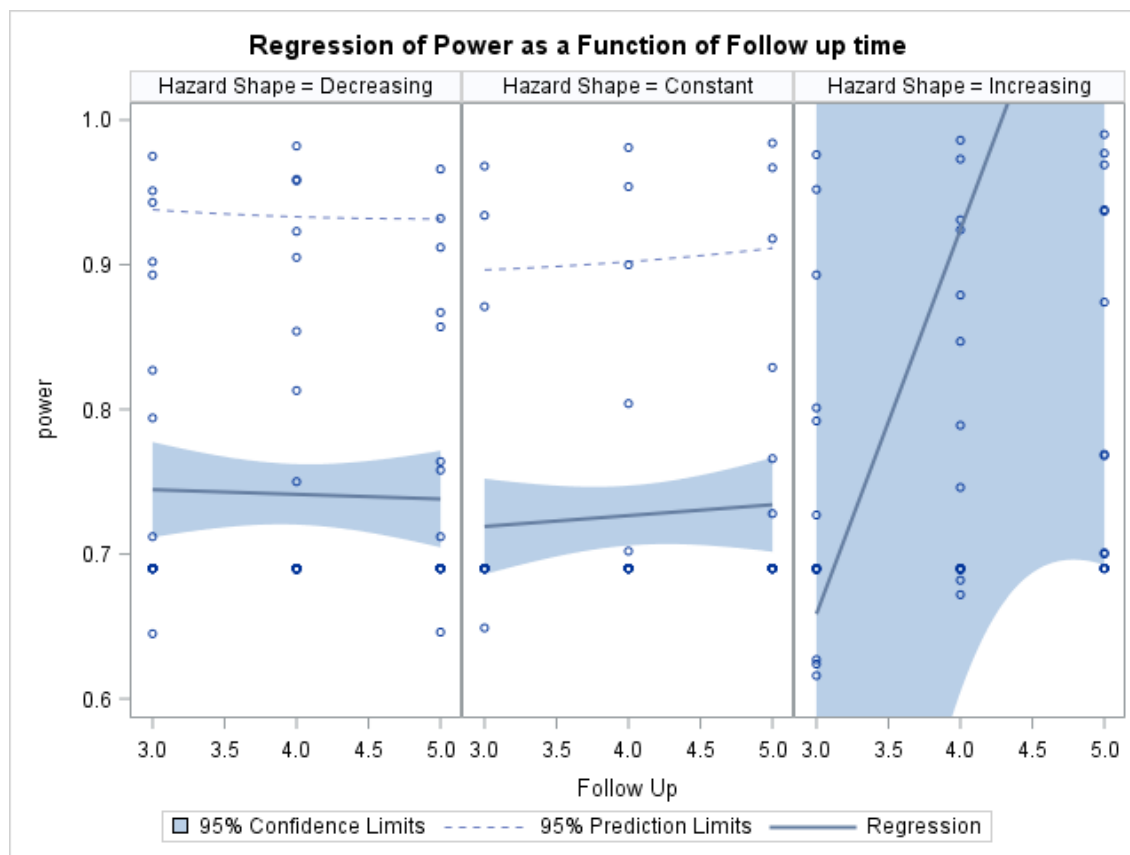


phenomena/disease in question. In clinical studies, it is often the case that hazard can be characterized based on the nature and severity of the disease. For example, in aggressive stages of breast cancer, the hazard of disease recurrence after mastectomy/surgery is expected to be increasing in time. This is in contrast to behavioral studies of recidivism after cessation of smoking, where resource allocation for the choice of follow-up periods, ranging from 3 to 5 years, as well as scale parameter values of 2.5 and 4.5. The latter impacts response variability, which in turn dictates clinical trials subjects' eligibility/selection criteria, thereby for overly homogeneous subjects, extends accrual periods beyond viably allowable resources.

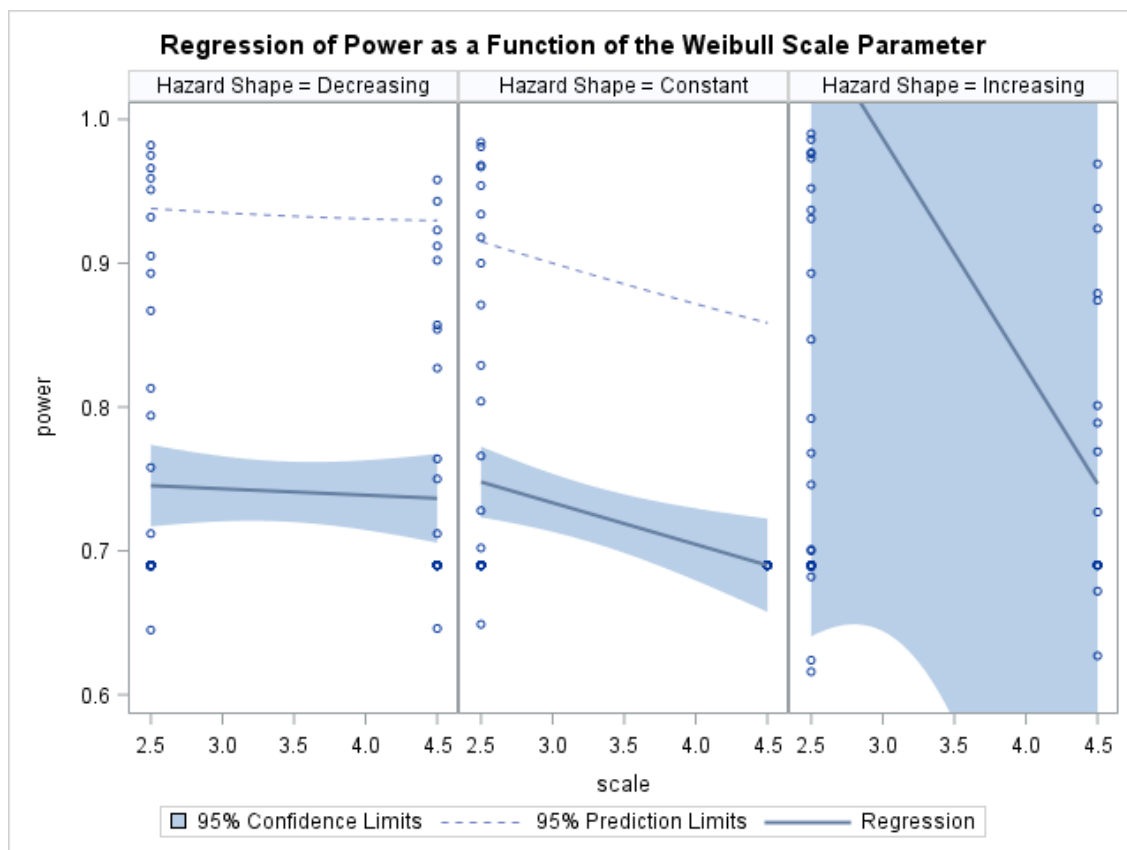
### Marginal effects of directional shifts in hazard function on the bottom edge of power function distribution

Although variable hazard seems to independently produce some tangible effects (variability) on the requirement of power and sample size, overall, the impact appears to be more pronounced at the *bottom* edge of the power function distribution. To that effect,

hazard is expected to be decreasing in time. The choice of HR values to produce alternative hypotheses was intended to mimic realistic study design scenarios, varying from 20% to 50% reduction in the tested group. Values less than 20% or greater than 50% may very well represent values of unworthy investigation or unrealistic expectation, respectively. We also appealed to logic in we completed a preliminary investigation of the behavior of the power function distribution between values of 0.7 and 0.8. It may be an oversimplification and of limited inference when applying regression techniques using one predictor, still, there seems to be some interaction between hazard function directional shape and change in each follow-up period and scale parameter. Note, in time-to-event studies, power is expected to be either monotonically non-decreasing or increasing, as a function of follow-up or scale/variability parameter, respectively. However, changes in power values in the range of 0.70 to 0.80 vis-à-vis increasing follow-up duration appear to be either minimally decreasing or increasing under decreasing or constant hazard, respectively and sharply increasing under increasing hazard function.



**Figure 6:** The converse is observed when evaluating changes in power as a function of scale parameters



**Figure 7:** It goes without saying that these observational findings require future study to substantiate.

This work is limited to two study groups (reference and experimental), though it can be easily extended to encompass multiple groups. Other scenarios of hazard shape can be investigated under more complex scenarios, where hazard function direction can change within the group and during follow-up. This work also (purposely) did not address the effects of non-uniform accrual, simply because our focus was on a one-time (final) data analysis scenario, where all subjects have essentially been recruited irrespective of their accrual patterns.

Furthermore, it is worth noting that the choice of sample size values, as well as shape parameter values, was discrete and the ultimate outcome of this work can (or should) be a frontend, user-friendly software that addresses these limitations and fully executes for infinite choices of hazard functions, sample sizes, and follow-up durations.

### Glossary

**Weibull distribution** is a mathematical formula describing the probability distribution of waiting time for an event to take place.

**Hazard function,  $h_x(x)$**  is the instantaneous rate at which an event under observation takes place given that it did not occur before up to time  $x$

**Null hypothesis** is an assumption conventionally assigned to the situation of no change in an outcome a) relative to a standard (or

baseline) state or b) between two or more independent groups

**Alternative hypothesis** is an assumption conventionally assigned to the complimenting situation under the null hypothesis. It pertains to existence of change in an outcome a) relative to a standard (or baseline) state or b) between two or more independent groups

**Type I error** is the likelihood of rejecting the null hypothesis when it is true. It is also known as the size of the test or significance level. It is usually preset before testing a hypothesis.

**Type II error** is the likelihood of rejecting the alternative hypothesis when it is true. Subtracting this error magnitude from the value “1” gives the statistical power of a test of hypothesis.

**Statistical power of a test** is the level of certainty in detecting preset change in an outcome a) relative to a standard state (or baseline) or b) between two or more independent groups. It is also determined/projected at the disgning stage of a research study.

### Acknowledgement:

This work was supported by 1) the NIH National Cancer Institute (P30CA013148, P50 CA098252, U24MD015970, U54-FOAP 310102-290001-21, U54CA280770, U54CA118938, U54CA118623, U54CA118948, UL1TR002378), 2) the Centers for Disease Control and Prevention (Grant # U48DP006411), 3) the National Center for Advancing Translational Sciences of the National Institutes of Health (Grant #UL1TR002378), 4) the

National Institute of Minority Health and Health Disparities of the National Institutes of Health under Award Number 1U24MD015970, 5) National Institute on Minority Health and Health Disparities (NIMHD, U24MD017138) and 6) National Research Mentoring Institute (U01GM132771). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies, including the NIH) or the CDC.

**Institutional Review Board Statement:** This work is methodological in nature and was developed using statistical methods, theories and simulations. Therefore, Institutional Review Board (or Ethics Committee) was not applicable.

**Informed Consent Statement:** There were no study participants associated with the secondary data analysis

**Data Availability Statement:** Data associated with this work has been simulated using specific statistical and mathematical structure and can be publicly accessible.

**Acknowledgments:** We acknowledge the Morehouse School of Medicine Prevention Researcher Center Research Development and Dissemination Group for their support in the conceptualization and implementation of this work.

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agency, NIH.

#### Abbreviations

**LRSSP:** Least Required Sample Size and Power

**HR:** Hazard Ratio

#### References

1. Chow S-C, Shao J, Wang H, Lokhnygina Y. Sample size calculations in clinical research, 3rd edition. Chapman & Hall/CRC Biostatistics Series. New York: Taylor & Francis; 2017.
2. Cox DR, Oakes D. Analysis of survival data. London UK: Chapman and Hall; 1984.
3. Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 1981;68:316–319.
4. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials*. 1981;2:93-113.
5. Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*. 1986;42:507-519.
6. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39:499-503.
7. Palta M, Amini SB. Consideration of covariates and stratification in sample size determination for survival time studies. *J Chronic Dis*. 1985;38:801-809.
8. Lakatos E. Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*. 1988;44:229–241.
9. Makuch RW, Simon RM. Sample size requirement for comparing time-to-failure among k treatment groups. *J Chronic Dis*. 1982;35:861-867.
10. Ahnn S, Anderson SJ. Sample size determination for comparing more than two survival distributions. *Stat Med*. 1995;14:2273-2282.
11. Machin D, Campbell MJ, Fayers PM, Pinol A. Sample size tables for clinical studies. Oxford: Blackwell; 1996.
12. Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Control Clin Trials* 2000;21:552–560.
13. Casteloe JM, O'Brien RG. Power and sample size determination for linear models. Paper 240-26: Proceedings of the Twenty-sixth Annual SAS Users Group International Conference; 2001; Cary, NC: SAS Institute Inc.
14. Lakatos E. Designing complex group sequential survival trials. *Stat Med*. 2002;21:1969-1989.
15. Shen Y, Cai J. Sample size reestimation for clinical trials with censored survival data. *J Am Stat Assoc*. 2003;98:418-426.
16. Song R, Kosorok MR, Cai J. Robust covariate-adjusted log-rank statistics and corresponding sample size formula for recurrent events data. *Biometrics*. 2008;64:741-750.
17. Zhang D, Quan H. Power and sample size calculation for log-rank test with a time lag in treatment effect. *Stat Med*. 2009;28:864-879.
18. Lachin JM. Sample size and power for a logrank test and Cox proportional hazards model with multiple groups and strata, or a quantitative covariate with multiple strata. *Stat Med*. 2013;32:4413-4425.
19. Chen LM, Ibrahim JG, Chu H. Sample size determination in shared frailty models for multivariate time-to-event data. *J Biopharm Stat*. 2014;24:908-923.
20. Heo M, Faith MS, Allison DB. Power and sample size for survival analysis under the Weibull distribution when the whole lifespan is of interest. *Mech Ageing Dev*. 1998;102:45–53.
21. Jeske DR, Flegal JM, Spindler S. Minimum size survival analysis sampling plans for comparing multiple treatment groups to a single control group. *Commun Stat Theory Methods*. 2014;43:2689-2701.
22. Wu J. Power and sample size for randomized phase III survival trials under the Weibull model. *J Biopharm Stat*. 2015;25:16-28.
23. Lim JH, Kim Y. Comparison of methods for estimation of sample sizes under the Weibull distribution. *Int J Appl Eng Res*. 2017;12:14273-14278.
24. Yang Y, Li W, Tang W, Li B, Zhang D. Sample sizes based on Weibull distribution and normal distribution for FRP tensile coupon test. *Materials (Basel)*. 2019 Jan;12:126. doi: 10.3390/ma12010126.