Title

> **stpower exponential** — Sample size and power for the exponential test

Syntax

Sample-size determination

  Specifying hazard rates

    <u>stpower</u> <u>expo</u>nential $\begin{bmatrix} h_1 \begin{bmatrix} h_2 \end{bmatrix} \end{bmatrix}$ $\begin{bmatrix} , \textit{ options} \end{bmatrix}$

  Specifying survival probabilities

    <u>stpower</u> <u>expo</u>nential $s_1 \begin{bmatrix} s_2 \end{bmatrix}$, t(*#*) $\begin{bmatrix} \textit{options} \end{bmatrix}$

Power determination

  Specifying hazard rates

    <u>stpower</u> <u>expo</u>nential $\begin{bmatrix} h_1 \begin{bmatrix} h_2 \end{bmatrix} \end{bmatrix}$, n(*numlist*) $\begin{bmatrix} \textit{options} \end{bmatrix}$

  Specifying survival probabilities

    <u>stpower</u> <u>expo</u>nential $s_1 \begin{bmatrix} s_2 \end{bmatrix}$, t(*#*) n(*numlist*) $\begin{bmatrix} \textit{options} \end{bmatrix}$

where

  h_1 is the hazard rate in the control group;

  h_2 is the hazard rate in the experimental group;

  s_1 is the survival probability in the control group at reference (base) time t; and

  s_2 is the survival probability in the experimental group at reference (base) time t. h_1, h_2 and s_1, s_2 may each be specified either as one number or as a list of values (see [U] **11.1.8 numlist**) enclosed in parentheses.

options	Description
Main	
t(*#*)	reference time t for survival probabilities s_1 and s_2
*alpha(*numlist*)	significance level; default is alpha(0.05)
*power(*numlist*)	power; default is power(0.8)
*beta(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	sample size; required to compute power
*hratio(*numlist*)	hazard ratio of the experimental group to the control group, h_2/h_1 or $\ln(s_2)/\ln(s_1)$; default is hratio(0.5)
onesided	one-sided test; default is two sided
*p1(*numlist*)	the proportion of subjects in the control group; default is p1(0.5), meaning equal group sizes
*nratio(*numlist*)	ratio of sample sizes, N_2/N_1; default is nratio(1), meaning equal group sizes
loghazard	power or sample-size computation for the test of the difference between log hazards; default is the test of the difference between hazards
unconditional	power or sample-size computation using the unconditional approach
parallel	treat number lists in starred options as parallel (do not enumerate all possible combinations of values) when multiple values per option are specified
Accrual/Follow-up	
fperiod(*#*)	length of the follow-up period; if not specified, the study is assumed to continue until all subjects experience an event (fail)
aperiod(*#*)	length of the accrual period; default is aperiod(0), meaning no accrual
aprob(*#*)	proportion of subjects accrued by time t^* under truncated exponential accrual; default is aprob(0.5)
aptime(*#*)	proportion of the accrual period, $t^*/$aperiod(), by which proportion of subjects in aprob() is accrued; default is aptime(0.5)
atime(*#*)	reference accrual time t^* by which the proportion of subjects in aprob() is accrued; default value is $0.5\times$aperiod()
ashape(*#*)	shape of the truncated exponential accrual distribution; default is ashape(0), meaning uniform accrual
lossprob(*# #*)	proportion of subjects lost to follow-up by time losstime() in the control and the experimental groups; default is lossprob(0 0), meaning no losses to follow-up
losstime(*#*)	(reference) time by which the proportion of subjects specified in lossprob() is lost to follow-up; default is losstime(1)
losshaz(*# #*)	loss hazard rates in the control and the experimental groups; default is losshaz(0 0), meaning no losses to follow-up

Reporting

<u>detail</u>	more detailed output
<u>table</u>	display results in a table with default columns
<u>col</u>umns(*colnames*)	display results in a table with specified *colnames* columns
<u>noti</u>tle	suppress table title
<u>nolegend</u>	suppress table legend
<u>colw</u>idth(# [# ...])	column widths; default is colwidth(9)
<u>sep</u>arator(#)	draw a horizontal separator line every # lines; default is separator(0), meaning no separator lines
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
<u>noheader</u>	suppress table header; seldom used
<u>continue</u>	draw a continuation border in the table output; seldom used

*Starred options may be specified either as one number or as a list of values (see [U] **11.1.8 numlist**).
noheader and continue are not shown in the dialog box.

colnames	Description
<u>a</u>lpha	significance level
<u>pow</u>er	power
<u>b</u>eta	type II error probability
n	total number of subjects
n1	number of subjects in the control group
n2	number of subjects in the experimental group
hr	hazard ratio
loghr	log of the hazard ratio (difference between log hazards)
diff	difference between hazards
h1	hazard rate in the control group
h2	hazard rate in the experimental group
s1	survival probability in the control group
s2	survival probability in the experimental group
t	reference survival time
p1	proportion of subjects in the control group
<u>nra</u>tio	ratio of sample sizes, experimental to control
<u>fp</u>eriod	follow-up period
<u>ap</u>eriod	accrual period
<u>apr</u>ob	% of subjects accrued by time atime (or by aptime % of accrual period)
<u>apt</u>ime	% of an accrual period by which aprob % of subjects are accrued
<u>at</u>ime	reference accrual time
<u>as</u>hape	shape of the accrual distribution
lpr1	proportion of subjects lost to follow-up in the control group
lpr2	proportion of subjects lost to follow-up in the experimental group
<u>los</u>stime	reference loss to follow-up time

lh1	loss hazard rate in the control group
lh2	loss hazard rate in the experimental group
eo	total expected number of events (failures) under the null
eo1	number of events in the control group under the null
eo2	number of events in the experimental group under the null
ea	total expected number of events (failures) under the alternative
ea1	number of events in the control group under the alternative
ea2	number of events in the experimental group under the alternative
lo	total expected number of losses to follow-up under the null
lo1	number of losses in the control group under the null
lo2	number of losses in the experimental group under the null
la	total expected number of losses to follow-up under the alternative
la1	number of losses in the control group under the alternative
la2	number of losses in the experimental group under the alternative

By default, the following *colnames* are displayed:

> power, n, n1, n2, and alpha are always displayed;
>
> h1 and h2 are displayed if hazard rates are specified, or s1 and s2 if survival probabilities are specified;
>
> diff if hazard difference test is specified, or loghr if log-hazard difference test is specified;
>
> aperiod if accrual period (aperiod()) is specified;
>
> fperiod if follow-up period (fperiod()) is specified; and
>
> lh1 and lh2 if losshaz() or lpr1 and lpr2 if lossprob() is specified.

Menu

Statistics > Survival analysis > Power and sample size

Description

stpower exponential estimates required sample size and power for survival analysis comparing two exponential survivor functions by using parametric tests for the difference between hazards or, optionally, for the difference between log hazards. It accommodates unequal allocation between the two groups, flexible accrual of subjects into the study, and group-specific losses to follow-up. The accrual distribution may be chosen to be uniform over the fixed accrual period, R, or truncated exponential over the period $[0, R]$. Losses to follow-up are assumed to be exponentially distributed. Also the computations may be carried out using the conditional or the unconditional approach.

You can use stpower exponential to

- calculate expected number of events and required sample size when you know power and effect size (supplied as hazard rates, survival probabilities, or hazard ratio) for studies with or without follow-up and accrual periods allowing for different accrual patterns and in the presence of losses to follow-up and

- calculate power when you know sample size and effect size (supplied as hazard rates, survival probabilities, or hazard ratio) for studies with or without follow-up and accrual periods allowing for different accrual patterns and in the presence of losses to follow-up.

If the t() option is specified, the command's input parameters are the values of survival probabilities in the control (or the less favorable) group, $S_1(t)$, and in the experimental group, $S_2(t)$, at a fixed time, t (reference survival time), specified in t(), given as s_1 and s_2, respectively. Otherwise, the

input parameters are assumed to be the values of the hazard rates in the control group, λ_1, and in the experimental group, λ_2, given as h_1 and h_2, respectively. If survival probabilities are specified, they are converted to hazard rates by using the formula for the exponential survivor function and the value of time t in t().

Options

    ┐ Main └

t(#) specifies a fixed time t (reference survival time) such that the proportions of subjects in the control and experimental groups still alive past this time point are as specified in s_1 and s_2. If this option is specified, the input parameters, s_1 and s_2, are the survival probabilities $S_1(t)$ and $S_2(t)$. Otherwise, the input parameters are assumed to be hazard rates, λ_1 and λ_2, given as h_1 and h_2, respectively.

alpha(*numlist*) sets the significance level of the test. The default is alpha(0.05).

power(*numlist*) sets the power of the test. The default is power(0.8). If beta() is specified, this value is set to be $1-$beta(). Only one of power() or beta() may be specified.

beta(*numlist*) sets the probability of a type II error of the test. The default is beta(0.2). If power() is specified, this value is set to be $1-$power(). Only one of beta() or power() may be specified.

n(*numlist*) specifies the number of subjects in the study to be used to compute the power of the test. By default, the sample-size calculation is assumed. This option may not be combined with beta() or power().

hratio(*numlist*) specifies the hazard ratio of the experimental group to the control group. The default is hratio(0.5). This value defines the clinically significant improvement of the experimental procedure over the control desired to be detected by a test, with a certain power specified in power(). If h_1 and h_2 (or s_1 and s_2) are given, hratio() is not allowed and the hazard ratio is computed as h_2/h_1 (or $\ln(s_2)/\ln(s_1)$).

onesided indicates a one-sided test. The default is two sided.

p1(*numlist*) specifies the proportion of subjects in the control group. The default is p1(0.5), meaning equal allocation of subjects to the control and the experimental groups. Only one of p1() or nratio() may be specified.

nratio(*numlist*) specifies the sample-size ratio of the experimental group relative to the control group, N_2/N_1. The default is nratio(1), meaning equal allocation between the two groups. Only one of nratio() or p1() may be specified.

loghazard requests sample-size or power computation for the test of the difference between log hazards (or the test of the log of the hazard ratio). This option implies uniform accrual. By default, the test of the difference between hazards is assumed.

unconditional requests that the unconditional approach be used for sample-size or power computation; see *The conditional versus unconditional approaches* and *Methods and formulas* for details.

parallel reports results sequentially (in parallel) over the list of numbers supplied to options allowing *numlist*. By default, results are computed over all combinations of the number lists in the following order of nesting: alpha(); p1() or nratio(); list of hazard rates h_1 and h_2 or survival probabilities s_1 and s_2; hratio(); power() or beta(); and n(). This option requires that options with multiple values each contain the same number of elements.

⌐ Accrual/Follow-up ¬

`fperiod(#)` specifies the follow-up period of the study, f. By default, it is assumed that subjects are followed up until the last subject experiences an event (fails). The (minimal) follow-up period is defined as the length of the period after the recruitment of the last subject to the study until the end of the study. If T is the duration of a study and R is the length of an accrual period, then the follow-up period is $f = T - R$.

`aperiod(#)` specifies the accrual period, R, during which subjects are to be recruited into the study. The default is `aperiod(0)`, meaning no accrual.

`aprob(#)` specifies the proportion of subjects expected to be accrued by time t^* according to the truncated exponential distribution. The default is `aprob(0.5)`. This option is useful when the shape parameter is unknown but the proportion of accrued subjects at a certain time is known. `aprob()` is often used in conjunction with `aptime()` or `atime()`. This option may not be specified with `ashape()` or `loghazard` and requires specifying a nonzero accrual period in `aperiod()`.

`aptime(#)` specifies the proportion of the accrual period, t^*/R, by which the proportion of subjects specified in `aprob()` is expected to be accrued according to the truncated exponential distribution. The default is `aptime(0.5)`. This option may not be combined with `atime()`, `ashape()`, or `loghazard` and requires specifying a nonzero accrual period in `aperiod()`.

`atime(#)` specifies the time point t^*, reference accrual time, by which the proportion of subjects specified in `aprob()` is expected to be accrued according to the truncated exponential distribution. The default value is $0.5 \times R$. This option may not be combined with `aptime()`, `ashape()`, or `loghazard` and requires specifying a nonzero accrual period in `aperiod()`. The value in `atime()` may not exceed the value in `aperiod()`.

`ashape(#)` specifies the shape, γ, of the truncated exponential accrual distribution. The default is `ashape(0)`, meaning uniform accrual. This option is not allowed in conjunction with `loghazard` and requires specifying a nonzero accrual period in `aperiod()`.

`lossprob(# #)` specifies the proportion of subjects lost to follow-up by time `losstime()` in the control and the experimental groups, respectively. The default is `lossprob(0 0)`, meaning no losses to follow-up. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `losshaz()`.

`losstime(#)` specifies the time at which the proportion of subjects specified in `lossprob()` is lost to follow-up, also referred to as the reference loss to follow-up time. The default is `losstime(1)`. This option requires specifying `lossprob()`.

`losshaz(# #)` specifies exponential hazard rates of losses to follow-up, η_1 and η_2, in the control and the experimental groups, respectively. The default is `losshaz(0 0)`, meaning no losses to follow-up. This option requires specifying `aperiod()` or `fperiod()` and may not be combined with `lossprob()`.

⌐ Reporting ¬

`detail` displays more detailed output; the expected number of events (failures) and losses to follow-up under the null and alternative hypotheses are displayed. This option is not appropriate with tabular output.

`table` displays results in a tabular format and is implied if any number list contains more than one element. This option is useful if you are producing results one case at a time and wish to construct your own custom table by using a `forvalues` loop.

columns(*colnames*) specifies results in a table with specified *colnames* columns. The order of the
columns in the output table is the same as the order of *colnames* specified in columns(). Column
names in columns() must be space-separated.

notitle prevents the table title from displaying.

nolegend prevents the table legend from displaying and column headers from being marked.

colwidth(# [# . . .]) specifies column widths. The default is 9 for all columns. The number of
specified values may not exceed the number of columns in the table. A missing value (.) may be
specified for any column to indicate the default width (9). If fewer widths are specified than the
number of columns in the table, the last width specified is used for the remaining columns.

separator(#) specifies how often separator lines should be drawn between rows of the table. The
default is separator(0), meaning that no separator lines should be displayed.

saving(*filename*[, replace]) creates a Stata data file (.dta file) containing the table values
with variable names corresponding to the displayed *colnames*. replace specifies that *filename* be
overwritten if it exists. saving() is only appropriate with tabular output.

The following options are available with stpower exponential but are not shown in the dialog
box:

noheader prevents the table header from displaying. This option is useful when the command is
issued repeatedly, such as within a loop. noheader implies notitle.

continue draws a continuation border at the bottom of the table. This option is useful when the
command is issued repeatedly within a loop.

Remarks and examples   stata.com

Remarks are presented under the following headings:

> *Introduction*
> *Other ways of specifying the effect size*
> *Sample-size determination by using different approximations*
> *Sample-size determination in the presence of censoring*
> *Nonuniform accrual and exponential losses to follow-up*
> *The conditional versus unconditional approaches*
> *Link to the sample-size and power computation for the log-rank test*
> *Power determination*

Introduction

Let $S_1(t)$ and $S_2(t)$ be the exponential survivor functions with hazard rates λ_1 and λ_2 in the control
and experimental groups, respectively. Define ψ to be the treatment effect that can be expressed as a
difference, $\delta = \lambda_2 - \lambda_1$, between hazard rates or as the log of the hazard ratio (a difference between
log hazard-rates), $\ln(\Delta) = \ln(\lambda_2/\lambda_1) = \ln(\lambda_2) - \ln(\lambda_1)$. Negative values of the treatment effect ψ
imply the superiority of the experimental treatment over the standard (control) treatment. Denote R
and T to be the length of the accrual period and the total duration of the study, respectively. Then,
the follow-up period f is $f = T - R$.

Consider a study designed to compare the exponential survivor functions, $S_1(t) = e^{-\lambda_1 t}$ and
$S_2(t) = e^{-\lambda_2 t}$, of the two treatment groups. The disparity in survivor functions may be tested using
the hazards λ_1 and λ_2 for the exponential model. Depending on the definition of the treatment effect
ψ, two test statistics based on the difference and on the log ratio of the hazards may be used to
conduct tests of the difference between survivor functions using respective null hypotheses, $H_0: \delta = 0$
and $H_0: \ln(\Delta) = 0$.

The basic formula for the sample-size and power calculations for the test of H_0: $\delta = 0$ is proposed by Lachin (1981). He also derives the equation relating the sample size and power allowing for uniform accrual of subjects into the study over the period from 0 to R. Lachin and Foulkes (1986) extend this formula to truncated exponential accrual over the interval 0 to R and exponential losses to follow-up over the interval 0 to T.

The simplest method for the sample-size and power calculations for the test of H_0: $\ln(\Delta) = 0$ is presented by George and Desu (1974). Rubinstein, Gail, and Santner (1981) extend their method to account for uniform accrual and exponential losses to follow-up and apply it to planning the duration of a survival study. The formula that relates the sample size and power for this test and takes into account the uniform accrual and exponential losses to follow-up is formulated by Lakatos and Lan (1992), based on the derivations of Rubinstein, Gail, and Santner (1981).

By default, `stpower exponential` computes the sample size required to achieve a specified power to detect a difference between hazard rates, $\delta_a = \lambda_{2a} - \lambda_{1a}$, using the method of Lachin (1981). If loghazard is specified, the sample size required to detect a log of the hazard ratio $\ln(\Delta_a) = \ln(\lambda_{2a}/\lambda_{1a})$ with specified power is reported using the formula derived by George and Desu (1974). In the presence of an accrual period, the methods of Lachin and Foulkes (1986) or for uniform accrual only, Rubinstein, Gail, and Santner (1981) (if loghazard and unconditional are specified), are used.

If power determination is desired, sample size `n()` must be specified. Otherwise, sample-size determination is assumed with `power(0.8)` (or, equivalently, `beta(0.2)`). The default setting for power or the probability of a type II error, a failure to reject the null hypothesis when the alternative hypothesis is true, may be changed by using `power()` or `beta()`, respectively.

The default probability of a type I error, a rejection of the null hypothesis when the null hypothesis is true, of a test is 0.05 but may be changed by using the `alpha()` option. One-sided tests may be requested by using the `onesided` option. The default equal-group allocation may be changed by specifying either `p1()` or `nratio()`.

By default, the estimates of sample sizes or power for the test of the difference between hazards are reported. This may be changed to the test versus the difference between log hazards by using the loghazard option. The default conditional approach may be replaced with the unconditional approach by using unconditional; see *The conditional versus unconditional approaches*.

If neither the length of a follow-up period, f, nor the length of an accrual period, R, is specified in `fperiod()` or `aperiod()`, respectively, the study is assumed to continue until all subjects experience an event (failure), regardless of how much time is required. If either of the two options is supplied, a study of length $T = R + f$ is assumed.

If an accrual period of length R is specified in the `aperiod()` option, uniform accrual over the period $[0, R]$ is assumed. The accrual distribution may be changed to truncated exponential when the shape parameter is specified in `ashape()`. The combination of the `aprob()` and `aptime()` (or `atime()`) options may be used in place of the `ashape()` option to request the desired shape of the truncated exponential accrual. To take into account exponential losses to follow-up, the `losshaz()` or `lossprob()` and `losstime()` options may be used. For examples, see *Nonuniform accrual and exponential losses to follow-up*.

Optionally, results may be displayed in a table by using `table` or `columns()` as demonstrated in *Sample-size determination in the presence of censoring* below and in *Creating output tables* of [ST] **stpower**. Refer to example 7 in *Power and effect-size determination* of [ST] **stpower logrank** and [ST] **stpower** to see how to obtain a graph of a power curve.

Other ways of specifying the effect size

Here we demonstrate how to provide the information about the disparity in the two exponential survivor functions, also known as the effect size, to `stpower exponential`. We separately consider the two tests, based on the hazard difference and based on the log of the hazard ratio (or on the log-hazard difference).

▷ Example 1: Effect size for the test on the difference between hazards

Consider a fictional study for which the goal is to compare the two exponential survivor functions of the control and the experimental groups, using the test based on the difference between hazard rates. The yearly hazard rate in the control group is known to be 0.4, corresponding to roughly 45% survival by 2 years. The effect size, expressed as the hazard ratio, of 0.5 corresponding to the reduction in hazard of the experimental group from 0.4 to 0.2 (or increase in survival to 67% by 2 years), is of interest. The investigators need an estimate of the required sample size to detect this effect size with 80% power with a two-sided, 0.05-level test, the default settings of `stpower exponential`. For simplicity, we also assume that there are enough resources to monitor the subjects until all of them fail (a type I study, as defined in the next section).

The hazard rate in the control group and the effect size are required in the computation. The value of the hazard rate in the control group may be supplied directly as argument h_1 (the first syntax) or computed using the supplied value of the survival probability, s_1, at the reference time specified in the `t()` option (the second syntax). The effect size may be supplied by using the `hratio()` option, or by directly supplying the hazard rate in the experimental group as argument h_2 or, if the `t()` option is specified, by supplying the survival probability in the experimental group as argument s_2 at reference time `t()`.

We demonstrate all the above by using our fictional example. To obtain the required sample size, we supply the hazard rate in the control group, 0.4, as argument h_1 and the hazard ratio of 0.5 in `hratio()`.

```
. stpower exponential 0.4, hratio(0.5)
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:

       alpha =    0.0500  (two sided)
          h1 =    0.4000
          h2 =    0.2000
       h2-h1 =   -0.2000
       power =    0.8000
          p1 =    0.5000

Estimated sample sizes:
           N =        74
          N1 =        37
          N2 =        37
```

Instead of the hazard rate, we specify the survival probability of 0.45 in the control group as argument s_2 and a reference survival time 2 in option `t()`.

```
. stpower exponential 0.45, t(2) hratio(0.5)
Note: input parameters are survival probabilities.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:
        alpha =    0.0500   (two sided)
           s1 =    0.4500
           s2 =    0.6708
            t =    2.0000
        h2-h1 =   -0.1996
        power =    0.8000
           p1 =    0.5000

Estimated sample sizes:
            N =         74
           N1 =         37
           N2 =         37
```

Because the hazard ratio of 0.5 is the default of hratio(), we could have omitted this option from the above syntaxes.

Now rather than specifying the effect size by using the hratio(0.5) option above, we supply the hazard rate in the experimental group, 0.2, as argument h_2,

```
. stpower exponential 0.4 0.2
```
(*output omitted*)

or the survival probability in the experimental group, 0.67, as argument s_2 and the reference survival time of 2 years in t(),

```
. stpower exponential 0.45 0.67, t(2)
```
(*output omitted*)

and obtain the same estimates as in the above output: 74 of the total sample and 37 of the group sample sizes.

◁

▷ Example 2: Effect size for the test on the log of the hazard ratio

Continuing with the above example, suppose that the two survivor functions are to be compared using the test based on the difference between the log hazards (or test based on the log of the hazard ratio). To request this test, the loghazard option must be specified.

Unlike the hazard-difference test, the log hazard-ratio test does not require the hazard rate in the control group for the computation when the study is continued until all subjects fail. The only required information is the size of the effect desired to be detected by this test.

To obtain the estimate of the sample size for the test based on the log of the hazard ratio, we specify the loghazard option and use the default value of the hazard ratio of 0.5:

```
. stpower exponential, loghazard

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log-hazard difference, conditional
Ho: ln(h2/h1) = 0

Input parameters:
        alpha =    0.0500   (two sided)
    ln(h2/h1) =   -0.6931
        power =    0.8000
           p1 =    0.5000

Estimated sample sizes:
            N =        66
           N1 =        33
           N2 =        33
```

Similarly to the examples for the hazard-difference test, the effect size may also be supplied via hazard rates,

```
. stpower exponential 0.4 0.2, loghazard
```
(*output omitted*)

or survival probabilities and reference survival time,

```
. stpower exponential 0.45 0.67, t(2) loghazard
```
(*output omitted*)

For a fixed-duration study, when either fperiod() or aperiod() is specified (see *Sample-size determination in the presence of censoring*), the sample-size computations for the log hazard-ratio test do require the value of the hazard rate in the control group.

◁

Sample-size determination by using different approximations

Consider the following two types of survival studies: the first type, a type I study, is when investigators have enough resources to monitor the subjects until all of them experience an event (failure) and the second type, a type II study, is when the study terminates after a fixed period of time, regardless of whether all subjects experienced an event by that time.

Here we explore sample-size estimates using different approximations for a type I study. Examples of sample-size determination for a type II study are presented in the next section.

In survival studies, the requirement for the sample size is based on the requirement to observe a certain number of events (failures) to ensure a prespecified power of a test to detect a difference in survivor functions. For a type I study, the number of subjects required for the study is the same as the number of events required to be observed in the study because all subjects experience an event by the end of the study.

▷ Example 3: Sample size using the Lachin method

Consider an example from Lachin (1981, 107). A clinical trial is to be conducted to compare the survivor functions in the control and the experimental groups with a one-sided exponential test, based on the difference between hazards, of the superiority of a new treatment (H_a: $\delta < 0$) for a disease with moderate levels of mortality. Subjects in the control group get a standard treatment and subjects in the experimental group receive a new treatment. From previous studies the yearly hazard rate for the standard treatment was found to be $\lambda_1 = 0.3$, corresponding to 50% survival after 2.3 years. The

investigators would like to know how many subjects are required to detect a reduction in hazard to $\lambda_2 = 0.2$ (H_a: $\delta = -0.1$), which corresponds to an increase in survival to 63% at 2.3 years, with 90% power, equal-sized groups, and a significance level, α, of 0.05.

To obtain the estimate of the sample size for the above study, we supply hazard rates 0.3 and 0.2 as arguments and specify the power(0.9) option for 90% power and the onesided option for a one-sided test.

```
. stpower exponential 0.3 0.2, power(0.9) onesided
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:

         alpha =    0.0500   (one sided)
            h1 =    0.3000
            h2 =    0.2000
         h2-h1 =   -0.1000
         power =    0.9000
            p1 =    0.5000

Estimated sample sizes:

             N =       218
            N1 =       109
            N2 =       109
```

From the output, a total of 218 events (subjects) must be observed (recruited) in a study to ensure a power of 90% of a one-sided exponential test to detect a 13% increase in survival probability of subjects in the experimental group with $\alpha = 0.05$. Our estimate of 218 of the total number of subjects (109 per group) required for the study is the same as the one reported in Lachin (1981, 107).

◁

▷ Example 4: Sample size using the George–Desu method

The syntax above reports the sample size obtained using the approximation of Lachin (1981) for the test based on the hazard difference. To obtain the sample size using the approximation of George and Desu (1974), for the equivalent alternative H_a: $\ln(\Delta) = -0.4055$ (a test based on the log of the hazard ratio), we need to specify the loghazard option.

```
. stpower exponential 0.3 0.2, power(0.9) onesided loghazard
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log-hazard difference, conditional
Ho: ln(h2/h1) = 0

Input parameters:

         alpha =    0.0500   (one sided)
            h1 =    0.3000
            h2 =    0.2000
      ln(h2/h1) =   -0.4055
         power =    0.9000
            p1 =    0.5000

Estimated sample sizes:

             N =       210
            N1 =       105
            N2 =       105
```

The George–Desu method yields a slightly smaller estimate, 210, of the total number of events (subjects). George and Desu (1974) studied the accuracy of the two approximations based on δ and $\ln(\Delta)$ and concluded that the former is slightly conservative; that is, it gives slightly larger sample-size estimates. The latter was found to be accurate to one or two units of the exact solution for equal-sized groups.

We could have also obtained the same results by using the `hratio()` option to specify the hazard ratio $0.2/0.3 = 0.66667$ instead of supplying arguments 0.3 and 0.2:

```
. stpower exponential, hratio(0.66667) power(0.9) onesided loghazard
```
   (*output omitted*)

◁

❑ Technical note

The above approach may also be used to obtain an approximation to the sample size or power for the exact F test of equality of two exponential mean analysis (life) times (using the relation between a mean and a hazard rate of the exponential distribution, $\mu = 1/\lambda$).

For example, the sample size of 210 obtained above may be used as an approximation to the number of subjects required in a study of which the goal is to detect an increase in a mean analysis (life) time of the experimental group from $3.33 = 1/0.3$ to $5 = 1/0.2$ by using the exact, one-sided, 0.05-level F test with 90% power.

The test statistic of the F test is a ratio of two sample means from two exponential distributions that has an exact F distribution. The George–Desu method is based on the normal approximation of the distribution of the log of this test statistic. George and Desu (1974) studied this approximation for equal-sized groups and some common values of significance levels, powers, and hazard ratios and found it to be accurate to one or two units of the exact solution.

❑

❑ Technical note

We can obtain the same results from `stpower exponential` as in the examples above by using the respective survival probabilities 0.5 and 0.63 at $t = 2.3$ as arguments in place of hazard rates 0.3 and 0.2. For example,

```
. stpower exponential 0.5 0.63, t(2.3) power(0.9) onesided
Note: input parameters are survival probabilities.
```
   (*output omitted*)

Other alternative syntaxes, using the `hratio()` option, are

```
. stpower exponential 0.3, hratio(0.66667) power(0.9) onesided
Note: input parameters are hazard rates.
```
   (*output omitted*)

where 0.3 is interpreted as a hazard rate because `t()` is not specified, and

```
. stpower exponential 0.5, t(2.3) hratio(0.6667) power(0.9) onesided
Note: input parameters are survival probabilities.
```
   (*output omitted*)

❑

Sample-size determination in the presence of censoring

Often in practice, investigators may not have enough resources to continue a study until all subjects experience an event and therefore plan to terminate the study after a fixed period, T. Some subjects may not experience an event by the end of the study, in which case the (administrative) censoring of subjects occurs. In the presence of censoring, the number of subjects required in a study will be larger than the number of events required to be observed in the study.

We investigate how terminating the study after some fixed period, T, before all subjects experience an event, affects the requirements for the sample size. The duration of a study is divided into two phases: an accrual phase of a length R, during which subjects are recruited to the study, and a follow-up phase of a length f, during which subjects are followed up until the end of the study and no new subjects enter the study. The duration of a study, T, is the sum of the lengths of the two phases.

Consider the following study designs. In the first study design, A, each subject is followed up for a length of time T. Here the minimum follow-up time f is equal to T, and, consequently, $R = 0$. In practice, however, subjects will often enter the study at random times and will be followed up until the end of a study at time T, in which case the subjects who come under observation later in time will have a shorter follow-up than subjects who entered the study at the beginning. Therefore, the minimum follow-up time f will be less than T, and R will be equal to $T - f$. Here the length of the accrual period, R, must be taken into account in the computations. In the presence of an accrual period, subjects may be recruited continuously during a period of length T ($R = T$, $f = 0$), the second study design, B. Or subjects may be recruited for a fixed period, R, and then followed up for a period of time, f, during which no new subjects enter the trial, so that the total duration of study is $T = R + f$, the third design, C.

▷ Example 5: Sample size in the presence of accrual and follow-up periods

Continuing example 3, assume that the investigators have resources to continue the study for only 5 years, $T = 5$. We tabulate sample-size values for different combinations of an accrual period, R, and a minimum follow-up period, f, such that $T = f + R = 5$. Because the `aperiod()` and `fperiod()` options allow specifying only one number, we can obtain the table by issuing `stpower exponential` repeatedly within a `forvalues` loop.

```
. local cont continue

. local columns columns(power n aperiod fperiod h1 h2 alpha)

. local T = 5

. forvalues R = 0/`T' {
  2.          if `R' == `T' {
  3.                  local cont
  4.          }
  5.          local f = `T' - `R'
  6.          stpower exponential 0.3 0.2, power(0.9) onesided
  >                  aperiod(`R') fperiod(`f') `header' `cont' `columns'
  7.          local header noheader
  8. }
Note: input parameters are hazard rates.
```

```
Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0
```

Power	N	AP+	FP	H1	H2	Alpha*
.9	304	0	5	.3	.2	.05
.9	322	1	4	.3	.2	.05
.9	344	2	3	.3	.2	.05
.9	378	3	2	.3	.2	.05
.9	426	4	1	.3	.2	.05
.9	502	5	0	.3	.2	.05

```
* one sided
+ uniform accrual; 50.00% accrued by 50.00% of AP
```

For each iteration of the loop, the value of a follow-up period, f, stored in macro f (see [P] **macro**), is computed using the current value of an accrual period, R, and the fixed-time study T, defined by macros R and T, respectively. The values 'R' and 'f' are then passed to the aperiod() and fperiod() options, respectively. We use the columns() option to obtain a customized table with columns power, n, aperiod, fperiod, h1, h2, and alpha. We store this option in local columns simply to avoid specifying a long expression in stpower exponential. We specify noheader in the syntax of stpower exponential each time it is called in the loop except for the first time, and we specify continue each time except for the final one.

The first and the last entries of the above table correspond to the extreme cases of no accrual (design A) and no follow-up (design B), respectively.

For a design A, the estimate of the sample size, 304, is larger than the earlier estimate of 218 from example 3. That is, if the study in example 3 terminates after 5 years, the requirement for the sample size increases by 39% to ensure that the same number of 218 events is observed.

By trying different values of the follow-up period, we may find that a 30-year follow-up is required if the investigators can recruit no more than 218 subjects: 30 years is required to observe all subjects to fail in this study.

```
. stpower exponential 0.3 0.2, power(0.9) onesided fperiod(30)
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:
        alpha =    0.0500  (one sided)
           h1 =    0.3000
           h2 =    0.2000
        h2-h1 =   -0.1000
        power =    0.9000
           p1 =    0.5000
 Accrual and follow-up information:
     duration =   30.0000
    follow-up =   30.0000

Estimated sample sizes:
            N =       218
           N1 =       109
           N2 =       109
```

For design B, instead of being monitored for 5 years, subjects are being continuously recruited during all of those 5 years; the total sample size increases from 304 to 502. The reason for such an

increase is that the average analysis time (the time when a subject is at risk of a failure) decreases from 5 to 2.5 and therefore reduces the probability of a subject failing by the end of the study.

In general, the estimates of the total sample size steadily increase as the length of the follow-up decreases. That is, the presence of a follow-up period reduces the requirement for the number of subjects in the study. For example, a clinical trial with a 3-year uniform accrual and a 2-year follow-up needs a total of 378 subjects (189 per group) compared with the total of 502 subjects required for a study with no follow-up and a 5-year accrual.

◁

## Nonuniform accrual and exponential losses to follow-up

In the presence of an accrual period, stpower exponential performs computations assuming uniform accrual over the period of time R, specified in aperiod(). The assumption of uniform accrual may be relaxed by requesting a truncated exponential accrual over the interval 0 to R with shape γ as specified in ashape(#). If the estimate of the shape parameter, γ, is unavailable, an alternative is to specify the proportion of subjects, $G(t^*)$, expected to be recruited by a fixed time, t^* (or by a proportion of the accrual period R, t^*/R), in the aprob() and atime() (or aptime()) options, respectively. This information is used to find the corresponding γ by using

$$G(t^*) = \{1 - \exp(-\gamma t^*)\}/\{1 - \exp(-\gamma R)\}$$

Also see Cleves et al. (2010, sec. 16.2) for more information and see *Methods and formulas* for technical details.

▷ Example 6: Truncated exponential entry distribution

Continuing example 5, we investigate the influence of nonuniform accrual on the estimate of the sample size for a study with a 3-year accrual and a 2-year follow-up. Suppose that the recruitment of subjects to the study is slow for most of the accrual period and increases rapidly toward the end of the recruitment. Consider an extreme case of such an accrual corresponding to shape parameter -6. The graph of uniform and truncated exponential with shape -6 entry distributions over $[0, 3]$ is given below.
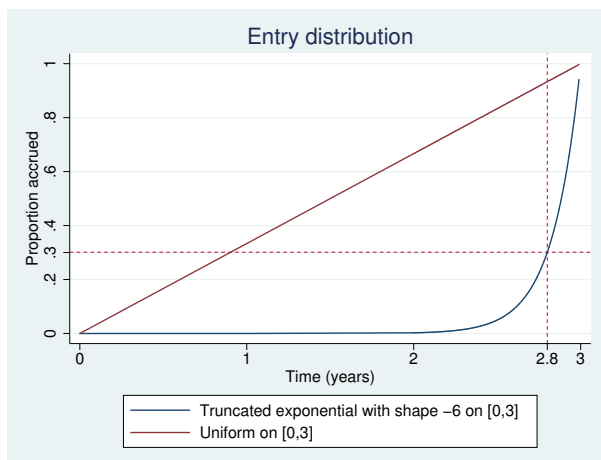
From the above graph, the accrual of subjects is extremely slow during most of the recruitment period, with 70% of subjects being recruited within the last few months of a 3-year accrual period. More precisely, according to the graph, 30% of subjects are expected to be recruited after 2.8 years.

To obtain the estimate of the sample size for this study, we type

```
. stpower exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> ashape(-6)
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:
         alpha =    0.0500  (one sided)
            h1 =    0.3000
            h2 =    0.2000
         h2-h1 =   -0.1000
         power =    0.9000
            p1 =    0.5000

 Accrual and follow-up information:
      duration =    5.0000
     follow-up =    2.0000
       accrual =    3.0000  (exponential)
    accrued(%) =     50.00  (by time t*)
            t* =    2.8845  (96.15% of accrual)

Estimated sample sizes:
             N =       516
            N1 =       258
            N2 =       258
```

and conclude that 516 subjects have to be recruited to this study. This sample size ensures 90% power of a one-sided, 0.05-level test to detect a reduction in hazard from 0.3 to 0.2 when the accrual of subjects follows the considered truncated exponential distribution. For this extreme case of a negative truncated exponential entry distribution (the concave entry distribution), the estimate of the sample 516 increases significantly compared with an estimate of 378 for a uniform entry distribution. On the other hand, a truncated exponential distribution with positive values of the shape parameter (convex entry distribution) will reduce the requirement for the sample size when compared with uniform accrual.

Suppose that we do not know (or do not wish to guess) the value of the shape parameter. The only information available to us from the above graph is that 30% of subjects are expected to be recruited after 2.8 years (or, equivalently, 50% of subjects are expected to be recruited by 96.15% of the accrual period from the previous output). We submit this information in the `aprob()` and `atime()` or `aprob()` and `aptime()` options, as shown below, and obtain the same estimate of the sample size 516.

```
. stpower exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> aprob(0.3) atime(2.8)
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:
        alpha =    0.0500   (one sided)
           h1 =    0.3000
           h2 =    0.2000
        h2-h1 =   -0.1000
        power =    0.9000
           p1 =    0.5000
 Accrual and follow-up information:
     duration =    5.0000
    follow-up =    2.0000
       accrual =    3.0000   (exponential)
   accrued(%) =      30.00   (by time t*)
           t* =    2.8000   (93.33% of accrual)
  Estimated sample sizes:
            N =        516
           N1 =        258
           N2 =        258
```

or, equivalently,

```
. stpower exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> aptime(0.9615)
    (output omitted )
```

Because 50% is the default value of `aprob()`, we omit this option in the above.

◁

Apart from administrative censoring, subjects may not experience an event by the end of the study because of being lost to follow-up for various reasons. See *Theory and terminology* in [ST] **stpower** and [ST] **Glossary** for a more detailed description. Rubinstein, Gail, and Santner (1981) and Lachin and Foulkes (1986) extend sample-size and power computations to take into account exponentially distributed losses to follow-up. In addition to being exponentially distributed, losses to follow-up are assumed to be independent of the survival times.

▷ Example 7: Exponential losses to follow-up

Suppose that in example 5, in the study with a 3-year uniform accrual and a 2-year follow-up, yearly loss hazards in the control and the experimental groups are 0.2. Groups may have different loss hazards, which can be specified in `losshaz()`.

```
. stpower exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> losshaz(0.2 0.2) detail
Note: input parameters are hazard rates.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, hazard difference, conditional
Ho: h2-h1 = 0

Input parameters:
          alpha =    0.0500  (one sided)
             h1 =    0.3000
             h2 =    0.2000
          h2-h1 =   -0.1000
          power =    0.9000
             p1 =    0.5000

 Accrual and follow-up information:
       duration =    5.0000
      follow-up =    2.0000
        accrual =    3.0000  (uniform)
            lh1 =    0.2000
            lh2 =    0.2000

Estimated sample sizes:
              N =       500
             N1 =       250
             N2 =       250

Estimated expected number of events:
         E|Ha =       213      E|Ho =       216
        E1|Ha =       121     E1|Ho =       108
        E2|Ha =        92     E2|Ho =       108

Estimated expected number of losses to follow-up:
         L|Ha =       173      L|Ho =       172
        L1|Ha =        81     L1|Ho =        86
        L2|Ha =        92     L2|Ho =        86
```

The sample size required for a one-sided, 0.05-level test to detect a reduction in hazard from 0.3 to 0.2 with 90% power increases from 378 (see example 5) to 500. We observe that for the extreme case of losses to follow-up, sample size increases significantly. A conservative adjustment commonly applied in practice is $N(1 + p_L)$, where p_L is the expected proportion of losses to follow-up in both groups combined. For this example, p_L may be computed as $0.5(0.369 + 0.324) \approx 0.35$ from table 2 of Lachin and Foulkes (1986). Then the conservative estimate of the sample size is $378(1 + 0.35) = 510$, which is slightly greater than 500, the actual required sample size.

We also requested that additional information about the expected number of events and losses to follow-up under the null and under the alternative hypothesis be displayed by using the detail option. From the above output, a total of 173 subjects (81 from the control and 92 from the experimental groups) are expected to be lost in the study with exponentially distributed losses with yearly rates of 0.2 in each group under the alternative hypothesis.

If the proportion of subjects lost to follow-up by a fixed period in each group is available rather than loss to follow-up rates, it can also be supplied by using the lossprob() and losstime() options. For example, in the above study approximately 33%, $1 - \exp(-0.2 \times 2) \approx 0.33$, of subjects in each group are lost at time 2 (years). We can obtain the same estimates of sample sizes with

```
. stpower exponential 0.3 0.2, power(0.9) onesided aperiod(3) fperiod(2)
> lossprob(0.33 0.33) losstime(2)
  (output omitted )
```

◁

The conditional versus unconditional approaches

Denote ψ to be the effect size, and $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ to be the maximum likelihood estimates of the respective hazard-rate parameters. Consider the two effect-size estimators based on the difference between the hazard rates, $\widehat{\lambda}_2 - \widehat{\lambda}_1$, and based on the log of the hazard ratio, $\ln(\widehat{\lambda}_2/\widehat{\lambda}_1)$. Both estimators are asymptotically normal under the null and under the alternative hypothesis.

We adopt Chow, Shao, and Wang (2008, 173) terminology when referring to the conditional and unconditional tests. The conditional test is the test that uses the constraint $\lambda_2 = \lambda_1$ (conditional on H_0) when computing the variance of the effect-size estimator under the null. The unconditional test is the test that does not use the above constraint when computing the variance of the effect-size estimator under the null. The score and the Wald tests are each one of the examples of conditional and unconditional tests, respectively. Chow, Shao, and Wang (2008) note that neither of the two tests (conditional or unconditional) is always more powerful than the other under the alternative hypothesis. Therefore, there is no definite recommendation of which one is preferable to be used in practice.

The conditional approach relies on the following relationship between sample size and power (given in Lachin 1981) to compute estimates of required sample size or power,

$$|\psi| = z_{1-\alpha} \left\{ \mathrm{Var}(\psi, H_0) \right\}^{1/2} + z_{1-\beta} \left\{ \mathrm{Var}(\psi, H_a) \right\}^{1/2}$$

where $z_{1-\alpha}$ and $z_{1-\beta}$ are the $(1-\alpha)$th and the $(1-\beta)$th quantiles of the standard normal distribution, and $\mathrm{Var}(\psi, H_0)$ and $\mathrm{Var}(\psi, H_a)$ are the asymptotic variances under the null and under the alternative, respectively, of the effect-size estimator, $\widehat{\psi}$. This approach uses the variance of the estimator conditional on the hypothesis type.

The unconditional approach replaces $\mathrm{Var}(\psi, H_0)$ with $\mathrm{Var}(\psi, H_a)$ in the above and uses the variance under the alternative to compute the estimates of sample size and power:

$$|\psi| = (z_{1-\alpha} + z_{1-\beta}) \left\{ \mathrm{Var}(\psi, H_a) \right\}^{1/2}$$

Therefore, the resulting formulas based on the two approaches are different.

Lakatos and Lan (1992) formulate the sample-size formula for the test of the log of the hazard ratio based on the method of Rubinstein, Gail, and Santner (1981). This formula is based on the unconditional approach. Lachin and Foulkes (1986) provide the sample-size formula for the test of the log of the hazard ratio that uses the conditional approach. They also present both conditional and unconditional versions of formulas for the test based on the difference between hazards. As noted by Lachin and Foulkes (1986), sample sizes estimated based on the unconditional approach will be larger than the estimates based on the conditional approach for equal-sized groups.

Both approaches are available with stpower exponential; the conditional is the default and the unconditional may be requested by specifying the unconditional option. Refer to *Methods and formulas* for the formulas underlying these approaches.

▷ Example 8: Sample size using the Rubinstein–Gail–Santner method

Consider the following scenario in Lakatos and Lan (1992, table I). A 10-year survival study with a 1-year accrual period and a 9-year follow-up is conducted to compare the survivor functions of the two groups by using a two-sided, 0.05 exponential test based on the log of the hazard ratio. The probability of surviving to the end of a study for subjects in the control group is 0.8 ($S_1(t) = 0.8$, $t = 10$). Subjects are recruited uniformly over the interval $[0, 1]$. Lakatos and Lan (1992) report an estimate of 664 for the sample size required to detect a change in the hazard of the experimental group

corresponding to the hazard ratio $\Delta = 0.5$ with 90% power by using the Rubinstein–Gail–Santner method. To obtain the estimates according to this method, we need to specify both `loghazard` and `unconditional`.

```
. stpower exponential 0.8, t(10) power(0.9) aperiod(1) fperiod(9) loghazard
> unconditional
Note: input parameters are survival probabilities.

Estimated sample sizes for two-sample comparison of survivor functions
Exponential test, log-hazard difference, unconditional
Ho: ln(h2/h1) = 0

Input parameters:
        alpha =    0.0500  (two sided)
           s1 =    0.8000
           s2 =    0.8944
            t =   10.0000
    ln(h2/h1) =   -0.6931
        power =    0.9000
           p1 =    0.5000
 Accrual and follow-up information:
     duration =   10.0000
    follow-up =    9.0000
      accrual =    1.0000  (uniform)
Estimated sample sizes:
            N =       664
           N1 =       332
           N2 =       332
```

◁

Because the default value of the hazard ratio is 0.5, we omit the `hratio(0.5)` option in the above. From the output, we obtain the same estimate of 664 of the sample size as reported in Lakatos and Lan (1992).

❑ Technical note

In the absence of censoring, the estimates of the sample size or power based on the test of log of the hazard ratio are the same for the conditional and the unconditional approaches. For example, both

```
. stpower exponential 0.8, t(10) power(0.9) loghazard
  (output omitted )
```

and

```
. stpower exponential 0.8, t(10) power(0.9) loghazard unconditional
  (output omitted )
```

produce the same estimate of the sample size (88). The asymptotic variance of maximum likelihood estimates of the log of the hazard ratio does not depend on hazard rates when there is no censoring and therefore does not depend on the type of hypothesis, $\text{Var}(\widehat{\psi}, H_0) = \text{Var}(\widehat{\psi}, H_a) = 2/N$.

❑

Although all the above examples demonstrated sample-size computation for equal-sized groups, unequal allocation may be taken into account by using the `p1()` or `nratio()` option.

Link to the sample-size and power computation for the log-rank test

▷ Example 9: Sample size using the Freedman and the Schoenfeld methods

Continuing examples 3 and 4, Lachin (1981, 106) gives another approximation to obtain the estimate of the sample size under the equal-group allocation. This approximation coincides with the formula derived by Freedman (1982) for the number of events in the context of the log-rank test. We can obtain such an estimate by using `stpower logrank` and by specifying the hazard ratio computed earlier of 0.66667:

```
. stpower logrank, hratio(0.66667) power(0.9) onesided

Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
Ho: S1(t) = S2(t)

Input parameters:
      alpha =    0.0500   (one sided)
     hratio =    0.6667
      power =    0.9000
         p1 =    0.5000

Estimated number of events and sample sizes:
          E =       216
          N =       216
         N1 =       108
         N2 =       108
```

The estimate, 216, of the sample size is the same as given in Lachin (1981, 107) and is slightly smaller than the estimate, 218, obtained in example 3, and larger than 210, obtained using the George–Desu method in example 4.

The approximation due to George and Desu (1974) is the same as the approximation to the number of events derived by Schoenfeld (1981) in application to the log-rank test. We can confirm that by using

```
. stpower logrank, hratio(0.66667) power(0.9) onesided schoenfeld
  (output omitted)
```

and obtain the same estimate of 210 as using `stpower exponential` with the `loghazard` option in example 4.

◁

Power determination

Power determination may be requested by specifying the `n()` option. We verify the power computation for the study from example 6. We expect the power estimate to be close to 0.9.

▷ Example 10: Power determination

The only thing we change in the syntax of `stpower exponential` from example 6 is replacing the `power(0.9)` option with the `n(664)` option:

```
. stpower exponential 0.8, t(10) n(664) aperiod(1) fperiod(9) loghazard
> unconditional
Note: input parameters are survival probabilities.

Estimated power for two-sample comparison of survivor functions
Exponential test, log-hazard difference, unconditional
Ho: ln(h2/h1) = 0

Input parameters:
        alpha =     0.0500   (two sided)
           s1 =     0.8000
           s2 =     0.8944
            t =    10.0000
    ln(h2/h1) =    -0.6931
            N =        664
           p1 =     0.5000
 Accrual and follow-up information:
     duration =    10.0000
    follow-up =     9.0000
      accrual =     1.0000   (uniform)
Estimated power:
        power =     0.9000
```

We obtain the estimate of power 0.9.

◁

For examples of how such information may be used to construct power curves, see [ST] **stpower** and example 7 in *Power and effect-size determination* of [ST] **stpower logrank**.

▷ Example 11: Using streg to perform the log hazard-ratio test

In this example, we demonstrate the importance of sample-size computations to ensure a high power of a test to detect a difference between exponential survivor functions. We consider an asymptotic Wald (or normal z) test to test the equality of the log of the hazard ratio to zero.

Continuing example 8, suppose that the investigators have only 100 subjects available for the study. As we see below, the power to detect a 50% risk reduction (the hazard ratio of 0.5) of the experimental group reduces from 90% to 24%:

```
. stpower exponential 0.8, t(10) n(100) aperiod(1) fperiod(9) loghazard
> unconditional
Note: input parameters are survival probabilities.

Estimated power for two-sample comparison of survivor functions
Exponential test, log-hazard difference, unconditional
Ho: ln(h2/h1) = 0

Input parameters:
        alpha =     0.0500   (two sided)
           s1 =     0.8000
           s2 =     0.8944
            t =    10.0000
    ln(h2/h1) =    -0.6931
            N =        100
           p1 =     0.5000
 Accrual and follow-up information:
     duration =    10.0000
    follow-up =     9.0000
      accrual =     1.0000   (uniform)
Estimated power:
        power =     0.2414
```

To demonstrate the implication of this reduction, consider the following example. We generate the data according to the study from example 8 with the following code.

```
program simdata
        args n h1 h2 R
        set obs 'n'
        generate double entry = 'R'*runiform()
        generate double u = runiform()
        /* random allocation to two groups of equal sizes */
        generate double u1 = runiform()
        generate double u2 = runiform()
        sort u1 u2, stable
        generate byte drug = (_n<='n'/2)
        /* exponential failure times with rates h1 and h2 */
        generate double failtime = entry - ln(1-u)/'h1' if drug==0
        replace failtime = entry - ln(1-u)/'h2' if drug==1
end

. clear

. set seed 234

. quietly simdata 100 0.0223 0.0112 1
```

The entry times of subjects are generated from a uniform $[0, 1)$ distribution and stored in variable `entry`. The subjects are randomized to two groups of equal size of 50 subjects each. The survival times are generated from exponential distribution with the hazard rate of $-\ln(0.8)/10 = 0.0223$ in the control group, `drug` $= 0$, and the hazard rate of $0.5 \times 0.0223 = 0.0112$ in the experimental group, `drug` $= 1$, conditional on subjects' entry times in `entry`.

Before analyzing these survival data, we need to set it up properly using `stset`. The failure-time variable is `failtime`. The study terminates at $t = 10$, so we use `exit(time 10)` with `stset` to specify that all failure times past 10 are to be treated as censored. Because subjects enter the study at random times (`entry`) and become at risk of a failure upon entering the study, we also specify the `origin(entry)` option to ensure that the analysis time is adjusted for the entry times. For more details, see [ST] **stset**.

```
. stset failtime, exit(time 10) origin(entry)

     failure event:  (assumed to fail at time=failtime)
obs. time interval:  (origin, failtime]
 exit on or before:  time 10
    t for analysis:  (time-origin)
            origin:  time entry
_____

       100  total observations
         0  exclusions
_____

       100  observations remaining, representing
        15  failures in single-record/single-failure data
  891.5574  total analysis time at risk and under observation
                                          at risk from t =          0
                                 earliest observed entry t =          0
                                    last observed exit t =    9.99908
```

To perform the log hazard-ratio test, we fit an exponential regression model on `drug` by using `streg` (see [ST] **streg**). We can express the log of the hazard ratio in terms of regression coefficients as follows: $\ln(\Delta) = \ln(\lambda_2/\lambda_1) = \ln\{\exp(\beta_0 + \beta_1)/\exp(\beta_0)\} = \beta_1$, where β_0 and β_1 are the estimated coefficients for the constant and `drug` in the regression model. Then the test of H_0: $\ln(\Delta) = 0$ may be rewritten in terms of a coefficient on `drug` as H_0: $\beta_1 = 0$. This test is part of the standard output after `streg`.

```
. streg drug, distribution(exponential) nohr

        failure _d:  1 (meaning all fail)
  analysis time _t:  (failtime-origin)
            origin:  time entry
  exit on or before:  time 10

Iteration 0:   log likelihood = -54.114272
Iteration 1:   log likelihood = -53.090859
Iteration 2:   log likelihood = -53.053454
Iteration 3:   log likelihood = -53.053406
Iteration 4:   log likelihood = -53.053406

Exponential regression -- log relative-hazard form

No. of subjects =           100                    Number of obs   =         100
No. of failures =            15
Time at risk    =  891.5574257
                                                   LR chi2(1)      =        2.12
Log likelihood  =   -53.053406                     Prob > chi2     =      0.1452
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
drug	-.7729235	.5477226	-1.41	0.158	-1.84644	.300593
_cons	-3.756554	.3162278	-11.88	0.000	-4.376349	-3.136759

From the output table above, the p-value for a two-sided test of the coefficient for `drug`, 0.158, is greater than 0.05. On that basis, we do not have evidence to reject the null hypothesis of no difference between the two exponential survivor functions and therefore make an incorrect decision because we simulated the data with different group hazard rates. If we were to repeat this, say, 100 times, using different datasets simulated according to the alternative H_a: $\ln(\Delta) = \ln(0.5) = -0.6931$ (see [R] **simulate**), for roughly 76 of them we would have failed to reject the null hypothesis of no difference (a type II error). Therefore, more subjects are required to be able to detect the log of the hazard ratio of -0.4055 in this study.

◁

▷ Example 12: Using results from streg to perform the Wald test of hazard difference

We obtain the power of the test based on the difference between hazards for the study in the above example (omit the loghazard option from the previous syntax).

```
. stpower exponential 0.8, t(10) n(100) aperiod(1) fperiod(9) unconditional
Note: input parameters are survival probabilities.

Estimated power for two-sample comparison of survivor functions
Exponential test, hazard difference, unconditional
Ho: h2-h1 = 0

Input parameters:
       alpha =   0.0500   (two sided)
          s1 =   0.8000
          s2 =   0.8944
           t =  10.0000
       h2-h1 =  -0.0112
           N =      100
          p1 =   0.5000

 Accrual and follow-up information:
    duration =  10.0000
   follow-up =   9.0000
     accrual =   1.0000   (uniform)

Estimated power:
       power =   0.2458
```

We obtain a power estimate of 0.2458, which is close to 0.2414 from example 9.

To test the difference between hazard rates by using the Wald test, we express this difference in terms of coefficients, $\lambda_2 - \lambda_1 = \exp(\beta_0)\{\exp(\beta_1) - 1\}$, and use testnl ([R] **testnl**) after streg to perform the nonlinear hypothesis test of H_0: $\exp(\beta_0)\{\exp(\beta_1) - 1\} = 0$.

```
. testnl exp(_b[_cons])*(exp(_b[drug])-1) = 0

  (1)  exp(_b[_cons])*(exp(_b[drug])-1) = 0

              chi2(1) =        2.03
          Prob > chi2 =      0.1540
```

We obtain the same conclusions from the Wald test based on the difference between hazards as in example 9. That is, based on the p-value of 0.1540 we fail to reject the null hypothesis of no difference between hazards of two groups (or miss the alternative H_a: $\delta = -0.0112$ corresponding to reduction in hazard from roughly 0.02 to 0.01) for the data from example 9.

◁

❑ Technical note

Often in practice, to test the disparity in two exponential survivor functions, the log-rank test is used instead of the hazard-difference test. Also the Wald (or the score) test from the Cox model is used instead of the exponential log hazard-ratio test. Refer to [ST] **sts test** and [ST] **stcox** for examples on how to perform these tests (also see [ST] **stpower logrank** and [ST] **stpower cox**).

❑

❑ Technical note

Sometimes the estimates of sample size and power obtained under the assumption of the exponential model are used as an approximation to the results used in a more general context of the log-rank test or the Cox proportional hazards model. Refer to Lachin (2011, 483–484) for the rationale behind this. Also see Lakatos and Lan (1992) for a discussion of the circumstances under which sample-size estimates obtained assuming the exponential model may be inaccurate when used with more general proportional hazards models.

❑

Stored results

stpower exponential stores the following in r():

Scalars
r(power)	power of test
r(alpha)	significance level of test
r(hratio)	hazard ratio
r(onesided)	1 if one-sided test, 0 otherwise
r(h1)	hazard in the control group (if specified)
r(h2)	hazard in the experimental group
r(t)	reference survival time (if t() is specified)
r(p1)	proportion of subjects in the control group
r(fperiod)	length of the follow-up period (if specified)
r(aperiod)	length of the accrual period (if specified)
r(ashape)	shape parameter (if aperiod() is specified)
r(lh1)	loss hazard in the control group (if specified)
r(lh2)	loss hazard in the experimental group (if specified)
r(lt)	reference loss to follow-up time (if losstime() is specified)

Macros
    r(method)   type of method (`hazard difference` or `log-hazard difference`)
    r(accrual)   type of entry distribution (`uniform` or `exponential`) (if requested)
    r(type)   type of approach (`conditional` or `unconditional`)
Matrices
    r(N) 1×3 matrix of required sample sizes
    r(Pr) 1×4 matrix of probabilities of an event (when computed)
    r(Ea) 1×3 matrix of expected number of events under the alternative (when computed)
    r(Eo) 1×3 matrix of expected number of events under the null (when computed)
    r(La) 1×3 matrix of expected number of losses under the alternative (when computed)
    r(Lo) 1×3 matrix of expected number of losses under the null (when computed)

Methods and formulas

In addition to the notation given in *Introduction*, denote N, N_1, and N_2 to be the total number of subjects required for the study, the number of subjects in the control group, and the number of subjects in the experimental group, respectively. Let $p_1 = N_1/N$ and $p_2 = N_2/N = 1 - p_1$ be the proportions of subjects allocated to the control and the experimental groups; γ be the shape parameter of the truncated exponential distribution with p.d.f. $g(z) = \gamma \exp(-\gamma z)/\{1 - \exp(-\gamma R)\}$, $0 \leq z \leq R$, $\gamma \neq 0$; η_1 and η_2 be the loss hazards in the control and the experimental groups; and $z_{(1-\alpha/k)}$ and $z_{(1-\beta)}$ be the $(1 - \alpha/k)$th and the $(1 - \beta)$th quantiles of the standard normal distribution, with $k = 1$ for the one-sided test and $k = 2$ for the two-sided test. Denote $\overline{\lambda} = p_1\lambda_1 + p_2\lambda_2$. Recall that the difference between hazards is denoted by $\delta = \lambda_2 - \lambda_1$ and the hazard ratio is denoted by $\Delta = \lambda_2/\lambda_1$.

If survival probabilities $S_1(t)$ and $S_2(t)$ at a fixed time t are specified rather than hazard rates, the hazard rates are computed as $\lambda_i = -\ln\{S_i(t)\}/t$, $i = 1, 2$. If loss to follow-up probabilities $L_1(t)$ and $L_2(t)$ at a fixed time t are given instead of loss to follow-up hazard rates, the loss hazard rates are computed as $\eta_i = -\ln\{1 - L_i(t)\}/t$, $i = 1, 2$.

All formulas below are derived under the assumption of exponential survival distributions with hazard rates in the control and the experimental groups λ_1 and λ_2, respectively, and rely on large-sample properties of the maximum likelihood estimates of λ_1 and λ_2.

Denote $\xi_o = \zeta(\overline{\lambda}, \gamma, \eta_1)p_1^{-1} + \zeta(\overline{\lambda}, \gamma, \eta_2)p_2^{-1}$ and $\xi_a = \zeta(\lambda_1, \gamma, \eta_1)p_1^{-1} + \zeta(\lambda_2, \gamma, \eta_2)p_2^{-1}$.

The formula for the sample-size calculation using the conditional approach is

$$N = \frac{\left(z_{1-\alpha/k}\xi_o^{1/2} + z_{1-\beta}\xi_a^{1/2}\right)^2}{\psi^2}$$

and using the unconditional approach is

$$N = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2 \xi_a}{\psi^2}$$

where $\zeta(\lambda, \gamma, \eta) = \lambda^2/p_E$ if $\psi = \delta$, $\zeta(\lambda, \gamma, \eta) = 1/p_E$ if $\psi = \ln(\Delta)$, and p_E is to be defined later. λ and η denote a failure hazard rate and a loss to follow-up hazard rate.

In the absence of censoring, the overall probability of an event (failure), p_E, is set to 1. Here the resulting formula for the sample size for the log hazard-ratio test depends only on the ratio of hazards and not on the individual group hazard rates. The resulting sample size formula for the test of the difference may also be rewritten as a function of the ratio of hazards only. Therefore, under no censoring, for a fixed value of the hazard ratio $\Delta = \lambda_2/\lambda_1$, the estimates of the sample size (or power) will be constant with respect to varying hazard rates λ_1 and λ_2.

In the presence of censoring, when each subject is followed up for a fixed period $f = T$,

$$p_{\text{E}} = p_{\text{E}}(\lambda, \eta) = \frac{\lambda}{\lambda + \eta}[1 - \exp\{-(\lambda + \eta)T\}]$$

In the presence of an accrual period, the probability of an event is defined as

$$p_{\text{E}} = p_{\text{E}}(\lambda, \eta) = \frac{\lambda}{(\lambda + \eta)}\left[1 - \frac{\exp\{-(\lambda + \eta)(T - R)\} - \exp\{-(\lambda + \eta)T\}}{(\lambda + \eta)R}\right]$$

or

$$p_{\text{E}} = p_{\text{E}}(\lambda, \gamma, \eta) = \frac{\lambda}{(\lambda + \eta)}\left(1 + \frac{\gamma\exp\{-(\lambda + \eta)T\}[1 - \exp\{(\lambda + \eta - \gamma)R\}]}{(\lambda + \eta - \gamma)\{1 - \exp(-\gamma R)\}}\right)$$

under uniform or truncated exponential accrual with shape γ over $[0, R]$, respectively. Uniform accrual is assumed for $|\gamma| < 10^{-6}$.

The formulas are obtained from Lachin (1981), Lachin and Foulkes (1986), and Lakatos and Lan (1992). To avoid division by 0 in the case $\lambda + \eta = \gamma$, the probability of an event is taken to be the limit of the above expression, $p_{\text{E}} = \lim_{\lambda + \eta -> \gamma} p_{\text{E}}(\lambda, \gamma, \eta)$.

The final estimate of the sample size is rounded up to the nearest even integer under the equal-group allocation and rounded up to the nearest integer otherwise. The number of subjects required to be recruited in each group is obtained as $N_1 = \pi_1 N$ and $N_2 = N - N_1$, where N_1 is rounded down to the nearest integer.

The expected number of events and losses to follow-up are computed as suggested by Lachin and Foulkes (1986). Under the null hypothesis,

$$E_{H_0} = N_1 p_{\text{E}}(\overline{\lambda}, \gamma, \eta_1) + N_2 p_{\text{E}}(\overline{\lambda}, \gamma, \eta_2)$$
$$L_{H_0} = N_1(\eta_1/\overline{\lambda})p_{\text{E}}(\overline{\lambda}, \gamma, \eta_1) + N_2(\eta_2/\overline{\lambda})p_{\text{E}}(\overline{\lambda}, \gamma, \eta_2)$$

and under the alternative hypothesis,

$$E_{H_a} = N_1 p_{\text{E}}(\lambda_1, \gamma, \eta_1) + N_2 p_{\text{E}}(\lambda_2, \gamma, \eta_2)$$
$$L_{H_a} = N_1(\eta_1/\lambda_1)p_{\text{E}}(\lambda_1, \gamma, \eta_1) + N_2(\eta_2/\lambda_2)p_{\text{E}}(\lambda_2, \gamma, \eta_2)$$

For unconditional tests, the expected number of events and losses to follow-up under the null is computed by setting $\overline{\lambda} = \lambda_1$. The estimates of the expected number of events and losses to follow-up in each group are rounded to the nearest integer.

To obtain the estimate of the power, $1 - \beta$, the formulas for the sample size are solved for $z_{(1-\beta)}$ and the normal cumulative distribution function is used to obtain the corresponding probability $1 - \beta$.

To obtain the unknown shape parameter, γ, of a truncated exponential entry distribution, an iterative procedure is used to solve the equation

$$\pi = G(t^*) = \frac{1 - \exp(-\gamma t^*)}{1 - \exp(-\gamma R)}$$

for a given proportion of subjects π recruited at a given time, t^*, for $t^* \in [0, R]$.

References

Chow, S.-C., J. Shao, and H. Wang. 2008. *Sample Size Calculations in Clinical Research.* 2nd ed. New York: Marcel Dekker.

Cleves, M. A., W. W. Gould, R. G. Gutierrez, and Y. V. Marchenko. 2010. *An Introduction to Survival Analysis Using Stata.* 3rd ed. College Station, TX: Stata Press.

Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.

George, S. L., and M. M. Desu. 1974. Planning the size and duration of a clinical trial studying the time to some critical event. *Journal of Chronic Diseases* 27: 15–24.

Lachin, J. M. 1981. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 2: 93–113.

——. 2011. *Biostatistical Methods: The Assessment of Relative Risks.* 2nd ed. Hoboken, NJ: Wiley.

Lachin, J. M., and M. A. Foulkes. 1986. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 42: 507–519.

Lakatos, E., and K. K. G. Lan. 1992. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine* 11: 179–191.

Rubinstein, L. V., M. H. Gail, and T. J. Santner. 1981. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *Journal of Chronic Diseases* 34: 469–479.

Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68: 316–319.

Also see [ST] **stpower** for more references.

Also see

[ST] **stpower** — Sample size, power, and effect size for survival analysis

[ST] **stpower cox** — Sample size, power, and effect size for the Cox proportional hazards model

[ST] **stpower logrank** — Sample size, power, and effect size for the log-rank test

[ST] **streg** — Parametric survival models

[ST] **Glossary**

[PSS] **power** — Power and sample-size analysis for hypothesis tests

[R] **test** — Test linear hypotheses after estimation