

**power logrank** — Power analysis for the log-rank test

|                             |                                      |                                |                                      |
|-----------------------------|--------------------------------------|--------------------------------|--------------------------------------|
| <a href="#">Description</a> | <a href="#">Quick start</a>          | <a href="#">Menu</a>           | <a href="#">Syntax</a>               |
| <a href="#">Options</a>     | <a href="#">Remarks and examples</a> | <a href="#">Stored results</a> | <a href="#">Methods and formulas</a> |
| <a href="#">References</a>  | <a href="#">Also see</a>             |                                |                                      |

## Description

`power logrank` computes sample size, power, or effect size for survival analysis comparing survivor functions in two groups by using the log-rank test. The results can be obtained using the Freedman or Schoenfeld approaches. Effect size can be expressed as a hazard ratio or as a log hazard-ratio. The command supports unbalanced designs, and provides options to account for administrative censoring, uniform accrual, and withdrawal of subjects from the study. For power and sample-size analysis in a cluster randomized design, see [PSS-2] [power logrank, cluster](#).

## Quick start

Sample size for the log-rank test of  $H_0: \Delta = 0$  versus  $H_a: \Delta \neq 0$  using the Freedman method for alternative hazard ratio  $\Delta_a = 0.76$  without censoring and with default power of 0.8 and significance level  $\alpha = 0.05$

```
power logrank, hratio(.76)
```

Same as above, but use Schoenfeld's method

```
power logrank, hratio(.76) schoenfeld
```

Sample size for censored design with survival probabilities  $surv_1 = 0.3$  and  $surv_2 = 0.4$

```
power logrank .3 .4
```

Same as above, specified as  $surv_1 = 0.3$  and hazard ratio of 0.76

```
power logrank .3, hratio(.76)
```

Same as above, but for hazard ratios of 0.65, 0.7, 0.75, and 0.8

```
power logrank .3, hratio(.65(.05).8)
```

Same as above, but show results in a graph of hazard ratio versus sample size

```
power logrank .3, hratio(.65(.05).8) graph
```

Sample size for a one-sided test with  $\alpha = 0.01$

```
power logrank .3, hratio(.76) onesided alpha(.01)
```

Sample size adjusted for 10% withdrawal from the study

```
power logrank .3, hratio(.76) wdprob(.1)
```

Power for a design with censoring and a sample size of 300

```
power logrank .3 .4, n(300)
```

Same as above, but specify twice as many observations in the experimental group

```
power logrank .3 .4, n(300) nratio(2)
```

Effect size for a design without censoring, sample size of 300, power of 0.8, and default  $\alpha = 0.05$

```
power logrank, n(300) power(.8)
```

Same as above, but for a censored design with control-group survival probability of 0.3

```
power logrank .3, n(300) power(.8)
```

## Menu

Statistics > Power, precision, and sample size

## Syntax

*Compute sample size*

```
power logrank [surv1 [surv2]] [, power(numlist) options]
```

*Compute power*

```
power logrank [surv1 [surv2]], n(numlist) [options]
```

*Compute effect size*

```
power logrank [surv1], n(numlist) power(numlist) [options]
```

where *surv*<sub>1</sub> is the survival probability in the control (reference) group at the end of the study  $t^*$  and *surv*<sub>2</sub> is the survival probability in the experimental (comparison) group at the end of the study  $t^*$ . *surv*<sub>1</sub> and *surv*<sub>2</sub> may each be specified either as one number or as a list of values in parentheses (see [U] [11.1.8 numlist](#)).

| <i>options</i>   | Description  |
|--|--|
| Main   |  |
| * <u>alpha</u> ( <i>numlist</i> )                                    | significance level; default is <code>alpha(0.05)</code>  |
| * <u>power</u> ( <i>numlist</i> )                                    | power; default is <code>power(0.8)</code>  |
| * <u>beta</u> ( <i>numlist</i> )                                     | probability of type II error; default is <code>beta(0.2)</code>  |
| * <u>n</u> ( <i>numlist</i> )  | total sample size; required to compute power or effect size  |
| * <u>n1</u> ( <i>numlist</i> )                                       | sample size of the control group   |
| * <u>n2</u> ( <i>numlist</i> )                                       | sample size of the experimental group  |
| * <u>nratio</u> ( <i>numlist</i> )                                   | ratio of sample sizes, $N2/N1$ ; default is <code>nratio(1)</code> , meaning equal group sizes   |
| <u>nfractional</u>   | allow fractional sample sizes  |
| * <u>hratio</u> ( <i>numlist</i> )                                   | hazard ratio of the experimental to the control group; default is <code>hratio(0.5)</code>   |
| * <u>lnhratio</u> ( <i>numlist</i> )                                 | log hazard-ratio of the experimental to the control group  |
| <u>schoenfeld</u>  | use the formula based on the log hazard-ratio in calculations; default is to use the formula based on the hazard ratio   |
| <u>effect</u> ( <i>effect</i> )                                      | specify the type of effect to display; default is method-specific  |
| <u>direction</u> ( <u>lower</u>   <u>upper</u> )                     | direction of the effect for effect-size determination; default is <code>direction(lower)</code> , which means that the postulated value of the parameter is smaller than the hypothesized value  |
| <u>onesided</u>  | one-sided test; default is two sided   |
| <u>parallel</u>  | treat number lists in starred options or in command arguments as parallel when multiple values per option or argument are specified (do not enumerate all possible combinations of values)   |
| Censoring  |  |
| <u>simpson</u> (# # #   <i>matname</i> )                             | survival probabilities in the control group at three specific time points to compute the probability of an event (failure), using Simpson's rule under uniform accrual   |
| <u>st1</u> ( <i>varname<sub>s</sub></i> <i>varname<sub>t</sub></i> ) | variables <i>varname<sub>s</sub></i> , containing survival probabilities in the control group, and <i>varname<sub>t</sub></i> , containing respective time points, to compute the probability of an event (failure), using numerical integration under uniform accrual |
| * <u>wdprob</u> ( <i>numlist</i> )                                   | proportion of subjects anticipated to withdraw from the study; default is <code>wdprob(0)</code>   |
| Table  |  |
| [ <u>no</u> ] <u>table</u> [ ( <i>tablespec</i> ) ]                  | suppress table or display results as a table; see [PSS-2] <a href="#">power</a> , <a href="#">table</a>  |
| <u>saving</u> ( <i>filename</i> [ , <code>replace</code> ])          | save the table data to <i>filename</i> ; use <code>replace</code> to overwrite existing <i>filename</i>  |
| Graph  |  |
| <u>graph</u> [ ( <i>graphopts</i> ) ]                                | graph results; see [PSS-2] <a href="#">power</a> , <a href="#">graph</a>   |

|                            |  |
|----------------------------|--|
| iteration                  |  |
| <code>init(#)</code>       | initial value for effect size  |
| <code>iterate(#)</code>    | maximum number of iterations; default is <code>iterate(500)</code>           |
| <code>tolerance(#)</code>  | parameter tolerance; default is <code>tolerance(1e-12)</code>                |
| <code>ftolerance(#)</code> | function tolerance; default is <code>ftolerance(1e-12)</code>                |
| <code>[no]log</code>       | suppress or display iteration log  |
| <code>[no]dots</code>      | suppress or display iterations as dots                                       |
| <code>cluster</code>       | perform computations for a CRD;<br>see [PSS-2] <b>power logrank, cluster</b> |
| <code>notitle</code>       | suppress the title   |

---

\*Specifying a list of values in at least two starred options, or at least two command arguments, or at least one starred option and one argument results in computations for all possible combinations of the values; see [U] 11.1.8 **numlist**. Also see the `parallel` option.

`collect` is allowed; see [U] 11.1.10 **Prefix commands**.

`cluster` and `notitle` do not appear in the dialog box.

| <i>effect</i>         | Description      |
|-----------------------|------------------|
| <code>hratio</code>   | hazard ratio     |
| <code>lnhratio</code> | log hazard-ratio |

---

where *tablespec* is

`column[:label] [column[:label] [...]] [, tableopts]`

*column* is one of the columns defined below, and *label* is a column label (may contain quotes and compound quotes).

| <i>column</i>         | Description  | Symbol        |
|-----------------------|--|---------------|
| <code>alpha</code>    | significance level   | $\alpha$      |
| <code>power</code>    | power  | $1 - \beta$   |
| <code>beta</code>     | type II error probability                                      | $\beta$       |
| <code>N</code>        | total number of subjects                                       | $N$           |
| <code>N1</code>       | number of subjects in the control group                        | $N_1$         |
| <code>N2</code>       | number of subjects in the experimental group                   | $N_2$         |
| <code>nratio</code>   | ratio of sample sizes, experimental to control                 | $N_2/N_1$     |
| <code>delta</code>    | effect size  | $\delta$      |
| <code>E</code>        | total number of events (failures)                              | $E$           |
| <code>hratio</code>   | hazard ratio   | $\Delta$      |
| <code>lnhratio</code> | log hazard-ratio   | $\ln(\Delta)$ |
| <code>s1</code>       | survival probability in the control group                      | $S_1(T)$      |
| <code>s2</code>       | survival probability in the experimental group                 | $S_2(T)$      |
| <code>Pr_E</code>     | overall probability of an event (failure)                      | $p_E$         |
| <code>Pr_w</code>     | probability of withdrawals                                     | $p_w$         |
| <code>target</code>   | target parameter; <code>hratio</code> or <code>lnhratio</code> |               |
| <code>_all</code>     | display all supported columns                                  |               |

---

Column `beta` is shown in the default table in place of column `power` if specified.

Column `lnhratio` is shown in the default table in place of column `hratio` if specified.

Columns `s1` and `s2` are available only when specified.

Columns `nratio` and `Pr_w` are shown in the default table if specified.

## Options

### Main

`alpha()`, `power()`, `beta()`, `n()`, `n1()`, `n2()`, `nratio()`, `nfractional`; see [PSS-2] [power](#).

`hratio(numlist)` specifies the hazard ratio (effect size) of the experimental group to the control group. The default is `hratio(0.5)`. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by the log-rank test with a certain power.

You can specify an effect size either as a hazard ratio in `hratio()` or as a log hazard-ratio in `lnhratio()`. The default is `hratio(0.5)`. If both arguments `surv1` and `surv2` are specified, `hratio()` is not allowed and the hazard ratio is instead computed as  $\ln(\text{surv}_2)/\ln(\text{surv}_1)$ .

This option is not allowed with the effect-size determination and may not be combined with `lnhratio()`.

`lnhratio(numlist)` specifies the log hazard-ratio (effect size) of the experimental group to the control group. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by the log-rank test with a certain power.

You can specify an effect size either as a hazard ratio in `hratio()` or as a log hazard-ratio in `lnhratio()`. The default is `hratio(0.5)`. If both arguments `surv1` and `surv2` are specified, `lnhratio()` is not allowed and the log hazard-ratio is computed as  $\ln\{\ln(\text{surv}_2)/\ln(\text{surv}_1)\}$ .

This option is not allowed with the effect-size determination and may not be combined with `hratio()`.

`schoenfeld` requests calculations using the formula based on the log hazard-ratio, according to [Schoenfeld \(1981\)](#). The default is to use the formula based on the hazard ratio, according to [Freedman \(1982\)](#).

`effect(effect)` specifies the type of the effect size to be reported in the output as `delta`. *effect* is one of `hratio` or `lnhratio`. By default, the effect size `delta` is a hazard ratio, `effect(hratio)`, for a hazard-ratio test and a log hazard-ratio, `effect(lnhratio)`, for a log hazard-ratio test (when `schoenfeld` is specified).

`direction()`, `onesided`, `parallel`; see [PSS-2] [power](#). `direction(lower)` is the default.

### Censoring

`simpson(# # # |matname)` specifies survival probabilities in the control group at three specific time points to compute the probability of an event (failure) using Simpson's rule under the assumption of uniform accrual. Either the actual values or a  $1 \times 3$  matrix, *matname*, containing these values can be specified. By default, the probability of an event is approximated as an average of the failure probabilities  $1-s_1$  and  $1-s_2$ ; see [Methods and formulas](#). `simpson()` may not be combined with `st1()` and may not be used if command argument `surv1` or `surv2` is specified. This option is not allowed with effect-size computation.

`st1(varnames varnamet)` specifies variables *varname<sub>s</sub>*, containing survival probabilities in the control group, and *varname<sub>t</sub>*, containing respective time points, to compute the probability of an event (failure) using numerical integration under the assumption of uniform accrual; see [R] **dydx**. The minimum and the maximum values of *varname<sub>t</sub>* must be the length of the follow-up period and the duration of the study, respectively. By default, the probability of an event is approximated as an average of the failure probabilities  $1-s_1$  and  $1-s_2$ ; see *Methods and formulas*. `st1()` may not be combined with `simpson()` and may not be used if command argument *surv<sub>1</sub>* or *surv<sub>2</sub>* is specified. This option is not allowed with effect-size computation.

`wdprob(numlist)` specifies the proportion of subjects anticipated to withdraw from the study. The default is `wdprob(0)`. `wdprob()` is allowed only with sample-size computation.

## Table

`table`, `table()`, `notable`; see [PSS-2] **power, table**.

`saving()`; see [PSS-2] **power**.

## Graph

`graph`, `graph()`; see [PSS-2] **power, graph**. Also see the *column* table for a list of symbols used by the graphs.

## Iteration

`init(#)` specifies an initial value for the estimated hazard ratio or, if `schoenfeld` is specified, for the estimated log hazard-ratio during the effect-size determination.

`iterate()`, `tolerance()`, `ftolerance()`, `log`, `nolog`, `dots`, `nodots`; see [PSS-2] **power**.

The following options are available with `power logrank` but are not shown in the dialog box:

`cluster`; see [PSS-2] **power logrank, cluster**.

`notitle`; see [PSS-2] **power**.

## Remarks and examples

[stata.com](http://www.stata.com)

Remarks are presented under the following headings:

*Introduction*

*Using power logrank*

*Computing sample size*

*Computing sample size in the absence of censoring*

*Computing sample size in the presence of censoring*

*Withdrawal of subjects from the study*

*Including information about subject accrual*

*Computing power*

*Computing effect size*

*Testing a hypothesis about two survivor functions using the log-rank test*

This entry describes the `power logrank` command and the methodology for power and sample-size analysis for a two-sample comparison of survivor functions using the log-rank test. See [PSS-2] **Intro (power)** for a general introduction to power and sample-size analysis and [PSS-2] **power** for a general introduction to the `power` command using hypothesis tests. See *Survival data* in [PSS-2] **Intro (power)** for an introduction to power and sample-size analysis for survival data. For power and sample-size analysis in a cluster randomized design, see [PSS-2] **power logrank, cluster**.

## Introduction

Consider a survival study comparing the survivor functions in two groups using the log-rank test. Let  $S_1(t)$  and  $S_2(t)$  denote the survivor functions of the control and the experimental groups, respectively. The key assumption of the log-rank test is that the hazard functions are proportional. That is,  $h_2(t) = \Delta h_1(t)$  for any  $t$  or, equivalently,  $S_2(t) = \{S_1(t)\}^\Delta$ , where  $\Delta$  is the hazard ratio. If  $\Delta < 1$ , the survival in the experimental group is higher relative to the survival in the control group; the new treatment is superior to the standard treatment. If  $\Delta > 1$ , then the standard treatment is superior to the new treatment. Under the proportional-hazards assumption, the test of the equality of the two survivor functions  $H_0: S_1(t) = S_2(t)$  versus  $H_a: S_1(t) \neq S_2(t)$  is equivalent to the test  $H_0: \Delta = 1$  versus  $H_a: \Delta \neq 1$  or  $H_0: \ln(\Delta) = 0$  versus  $H_a: \ln(\Delta) \neq 0$ .

The methods implemented in `power logrank` for power and sample-size analysis relate the power of the log-rank test directly to the number of events observed in the study. Depending on whether censoring occurs in a study, the required number of subjects is either equal to the number of events or is computed using the estimates of the number of events and the combined probability of an event (failure). Thus, in the presence of censoring, in addition to the number of events, the probability of a subject not being censored (failing) needs to be estimated to obtain the final estimate of the required number of subjects in the study.

To determine the required number of events, the investigator must specify the size or significance level,  $\alpha$ , and the clinically significant difference between the two treatments (effect size) to be detected by the log-rank test,  $H_a: \Delta = \Delta_a$ , with prespecified power  $1 - \beta$ . The effect size, a difference between the two treatments, is usually expressed as a hazard ratio,  $\Delta_a$ , using the `hratio()` option. Alternatively, you may specify an effect size as a log hazard-ratio,  $\ln(\Delta_a)$ , in the `lnhratio()` option.

When all subjects fail by the end of the study (no censoring), a **type I study**, the information above is sufficient to obtain the number of subjects required in the study. Often, in practice, not all subjects fail by the end of the study, in which case censoring of subjects occurs (a **type II study**). Here the estimates of the survival probabilities in the control and experimental groups are necessary to estimate an overall probability of an event and, then, the required sample size.

`power logrank` supports two methods, those of [Freedman \(1982\)](#) and [Schoenfeld \(1981\)](#), to obtain the estimates of the number of events or power (see also [Marubini and Valsecchi \[1995, 127, 134\]](#) and [Collett \[2015, 473, 479\]](#)). The final estimates of the sample size are based on the approximation of the probability of an event due to [Freedman \(1982\)](#), the default, or in the presence of uniform accrual, due to [Schoenfeld \(1983\)](#) (see also [Collett 2015](#)).

You can use `power logrank` to

- compute required number of events and sample size when you know power and effect size (expressed as a hazard ratio or log hazard-ratio);
- compute power when you know sample size (number of events) and effect size (expressed as a hazard ratio or log hazard-ratio); or
- compute effect size (hazard ratio or log hazard-ratio) and experimental-group survival when you know sample size (number of events) and power.

You can also choose between the Freedman or Schoenfeld computational approaches, adjust results for administrative censoring, adjust results for uniform accrual of subjects to the study, and adjust results for withdrawal of subjects from the study.

## Using power logrank

`power logrank` computes sample size, power, or effect size for the log-rank test comparing the survivor functions in two groups. All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the `alpha()` option. You can specify the `onesided` option to request a one-sided test. By default, all computations assume a balanced- or equal-allocation design; see [PSS-4] **Unbalanced designs** for a description of how to specify an unbalanced design.

To compute a total sample size, you specify an effect size and optionally power of the test in the `power()` option. The default power is set to 0.8. By default, the computed sample size is rounded up. You can specify the `nfractional` option to see the corresponding fractional sample size; see *Fractional sample sizes* in [PSS-4] **Unbalanced designs** for an example. The `nfractional` option is allowed only for sample-size determination.

To compute power, you must specify the total sample size in the `n()` option and an effect size.

An effect size may be specified either as a hazard ratio supplied in the `hratio()` option or as a log hazard-ratio supplied in the `lnhratio()` option. If neither is specified, a hazard ratio of 0.5 is assumed.

To compute an effect size, which may be expressed either as a hazard ratio or as a log hazard-ratio, you must specify the total sample size in the `n()` option; the power in the `power()` option; and, optionally, the direction of the effect. The direction is lower by default, `direction(lower)`, which means that the target hazard ratio is assumed to be less than one or that target log hazard-ratio is negative. In other words, the experimental treatment is presumed to be an improvement over the control treatment. If you want to change the direction to upper, corresponding to the target hazard ratio being greater than one, use `direction(upper)`.

Instead of the total sample size `n()`, you can specify individual group sizes in `n1()` and `n2()`, or specify one of the group sizes and `nratio()` when computing power or effect size. Also see *Two samples* in [PSS-4] **Unbalanced designs** for more details.

As we mentioned earlier, the effect size for `power logrank` may be expressed as a hazard ratio or as a log hazard-ratio. By default, the effect size, which is labeled as `delta` in the output, corresponds to the hazard ratio for the Freedman method and to the log hazard-ratio for the Schoenfeld method. You can change this by specifying the `effect()` option: `effect(hratio)` (the default) reports the hazard ratio and `effect(lnhratio)` reports the log hazard-ratio.

By default, all computations assume no censoring. In the presence of **administrative censoring**, you must specify a survival probability at the end of the study in the control group as the first command argument. You can also specify a survival probability at the end of the study in the experimental group as the second command argument. Otherwise, it will be computed using the specified hazard ratio or log hazard-ratio and the control-group survival probability. To accommodate an **accrual period** under the assumption of uniform accrual, survival information may instead be supplied in option `simpson()` or in option `st1()`; see *Including information about subject accrual* for details.

When computing sample size, you can adjust for withdrawal of subjects from the study by specifying the anticipated proportion of withdrawals in the `wdprob()` option.

By default, `power logrank` performs computations for a hazard-ratio test. Use the `schoenfeld` option to request computations for a log-hazard-ratio test.

In the presence of censoring, effect-size determination requires iteration. The default initial value of the estimated hazard ratio or, if `schoenfeld` is specified, of log hazard-ratio is obtained based on the formula assuming no censoring. This value may be changed by specifying the `init()` option. See [PSS-2] **power** for the descriptions of other options that control the iteration procedure.



In the following sections, we describe the use of `power logrank` accompanied by examples for computing sample size, power, and effect size.

## Computing sample size

To compute sample size and number of events, you must specify an effect size (a hazard ratio or a log hazard-ratio) and, optionally, the power of the test in the `power()` option. A default power of 0.8 is assumed if `power()` is not specified. A hazard ratio of 0.5 is assumed if an effect size is not specified.

## Computing sample size in the absence of censoring

We demonstrate several examples of how to use `power logrank` to obtain the estimates of sample size and number of events using [Freedman \(1982\)](#) and [Schoenfeld \(1981\)](#) methods for uncensored data (a type I study when no censoring of subjects occurs).

### ► Example 1: Number of events (failures) using Freedman method

Consider a survival study to be conducted to compare the survivor function of subjects receiving a treatment (the experimental group) to the survivor function of those receiving a placebo or no treatment (the control group) using the log-rank test. Suppose that the study continues until all subjects fail (no censoring). The investigator wants to know how many events need to be observed in the study to achieve a power of 80% of a two-sided log-rank test with  $\alpha = 0.05$  to detect a 50% reduction in the hazard of the experimental group ( $\Delta_a = 0.5$ ). Because the default settings of `power logrank` are `power(0.8)`, `alpha(0.05)`, and `hratio(0.5)`, to obtain the estimate of the required number of events for the above study using the Freedman method (the default), we simply type

```
. power logrank
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
HO: HR = 1 versus Ha: HR != 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.5000 (hazard ratio)
      hratio =    0.5000

Censoring:
      Pr_E =    1.0000

Estimated number of events and sample sizes:
      E =        72
      N =        72
      N per group =    36
```

From the output, a total of 72 events (failures) must be observed to achieve the required power of 80%. Because all subjects experience an event by the end of the study (`Pr_E=1.0000`), the number of subjects required to be recruited to the study is equal to the number of events. That is, the investigator needs to recruit a total of 72 subjects (36 per group) to the study.

## ▷ Example 2: Number of events (failures) using Schoenfeld method

Following [example 1](#), we can request the Schoenfeld method by specifying the `schoenfeld` option.

```
. power logrank, schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
H0: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =   -0.6931 (log hazard-ratio)
      hratio =    0.5000
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample sizes:
      E =        66
      N =        66
      N per group =    33
```

We obtain a slightly smaller estimate (66) of the total number of events and subjects.



## □ Technical note

[Freedman \(1982\)](#) and [Schoenfeld \(1981\)](#) derive the formulas for the number of events based on the asymptotic distribution of the log-rank test statistic. [Freedman \(1982\)](#) uses the asymptotic mean and variance of the log-rank test statistic expressed as a function of the true hazard ratio,  $\Delta$ , whereas [Schoenfeld \(1981\)](#) (see also [Collett \[2015, 474\]](#)) bases the derivation on the asymptotic mean of the log-rank test statistic as a function of the true log hazard-ratio,  $\ln(\Delta)$ . We label the corresponding approaches as “Freedman method” and “Schoenfeld method” in the output.

For values of the hazard ratio close to one, the two methods tend to give similar results. In general, the Freedman method gives higher estimates than the Schoenfeld method. The performance of the Freedman method was studied by [Lakatos and Lan \(1992\)](#) and was found to slightly overestimate the sample size under the assumption of proportional hazards. [Hsieh \(1992\)](#) investigated the performance of the two methods under unequal allocation and concluded that Freedman’s formula predicts the highest power for the log-rank test when the sample-size ratio of the two groups equals the reciprocal of the hazard ratio. Schoenfeld’s formula predicts highest powers when sample sizes in the two groups are equal.



### ▷ Example 3: Unbalanced design

By default, `power logrank` computes sample size for a balanced- or equal-allocation design. If we know the allocation ratio of subjects between the groups, we can compute the required sample size and number of events for an unbalanced design by specifying the `nratio()` option.

Continuing with [example 1](#), we anticipate being able to recruit twice as many subjects in the experimental group; that is,  $n_2/n_1 = 2$ . We specify the `nratio(2)` option to compute the required sample size for the specified unbalanced design.

```
. power logrank, nratio(2)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
HO: HR = 1 versus Ha: HR != 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
      delta =    0.5000 (hazard ratio)
      hratio =    0.5000
      N2/N1 =    2.0000
Censoring:
      Pr_E =    1.0000
Estimated number of events and sample sizes:
      E =        63
      N =        63
      N1 =       21
      N2 =       42
```

We need a total of 63 subjects—21 in the control group and 42 in the experimental group.

Also see [Two samples in \[PSS-4\] Unbalanced designs](#) for more examples of unbalanced designs for two-sample tests.



### Computing sample size in the presence of censoring

Because of constraints on costs and time, it is often infeasible to continue the study until all subjects experience an event. Instead, the study terminates at some prespecified point in time. As a result, some subjects may not experience an event by the end of the study; that is, administrative censoring of subjects occurs. This increases the requirement on the number of subjects in the study to ensure that a certain number of events is observed.

In the presence of censoring (for a type II study), [Freedman \(1982\)](#) assumes the following. The analysis occurs at a fixed time  $t^*$  after the last patient was accrued, and all information about subject follow-up beyond time  $t^*$  is excluded. To minimize an overestimation of the sample size because of neglecting this information, the author suggests choosing  $t^*$  as the minimum follow-up time,  $f$ , beyond which the frequency of occurrence of events is low (the time at which, say, 85% of the total events expected are observed). Under this assumption, the number of required subjects does not depend on the rates of accrual and occurrence of events but only on the proportions of patients in the two treatment groups,  $s_1$  and  $s_2$ , surviving after  $f$ . See [Including information about subject accrual](#) about how to compute the sample size in the presence of a long accrual.

If censoring of subjects occurs, the probability of a subject not being censored needs to be estimated to obtain an accurate estimate of the required sample size. The assumption above justifies a simple procedure, suggested by [Freedman \(1982\)](#) and used by default by `power logrank`, to compute

this probability using the estimates of survival probabilities at the end of the study in the control and the experimental groups. Therefore, for a type II study (under administrative censoring), these probabilities must be supplied to `power logrank`.

#### ► Example 4: Sample size in the presence of censoring using Freedman method

Consider an example from [Machin et al. \(2009, 91\)](#) of a study of patients with resectable colon cancer. The goal of the study was to compare the efficacy of the drug levamisole against a placebo with respect to relapse-free survival, using a one-sided log-rank test with a significance level of 5%. The investigators anticipated a 10% increase (from 50% to 60%, with a respective hazard ratio of 0.737) in the survival of the experimental group with respect to the survival of the control (placebo) group at the end of the study. They wanted to detect this increase with a power of 80%. To obtain the required sample size, we enter the survival probabilities 0.5 and 0.6 as arguments and specify the `onesided` option to request a one-sided test.

```
. power logrank 0.5 0.6, onesided
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
      alpha = 0.0500
      power = 0.8000
      delta = 0.7370 (hazard ratio)
      hratio = 0.7370
Censoring:
      s1 = 0.5000
      s2 = 0.6000
      Pr_E = 0.4500
Estimated number of events and sample sizes:
      E = 270
      N = 600
      N per group = 300
```

From the above output, the investigators would have to observe a total of 270 events (relapses) to detect a 26% decrease in the hazard ( $\Delta_a = 0.737$ ) of the experimental group relative to the hazard of the control group with a power of 80% using a one-sided log-rank test with  $\alpha = 0.05$ . They would have to recruit a total of 600 patients (300 per group) to observe that many events.

In contrast, in the absence of censoring, only 270 subjects would have been required to detect a decrease in hazard corresponding to  $\Delta_a = 0.737$ :

```

. power logrank, hratio(0.737) onesided
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = 0.7370 (hazard ratio)
    hratio = 0.7370
Censoring:
    Pr_E = 1.0000
Estimated number of events and sample sizes:
    E = 270
    N = 270
    N per group = 135

```

◀

### ► Example 5: Sample size in the presence of censoring using Schoenfeld method

If we wanted to use the Schoenfeld method to calculate sample size for the study described in [example 4](#), we could type

```

. power logrank 0.5 0.6, onesided schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
H0: ln(HR) = 0 versus Ha: ln(HR) < 0
Study parameters:
    alpha = 0.0500
    power = 0.8000
    delta = -0.3052 (log hazard-ratio)
    hratio = 0.7370
Censoring:
    s1 = 0.5000
    s2 = 0.6000
    Pr_E = 0.4500
Estimated number of events and sample sizes:
    E = 266
    N = 590
    N per group = 295

```

We find that 590 subjects are required in the study to observe a total of 266 events to ensure a power of 80%.

See the technical note in [Computing sample size in the absence of censoring](#) for a brief comparison of the Freedman and Schoenfeld methods.

◀

## ▷ Example 6: Alternative ways of specifying effect

If we wish, we can redefine effect size  $\delta$  in [example 4](#) to be a log hazard-ratio by specifying the `effect()` option.

```
. power logrank 0.5 0.6, onesided effect(lnhratio)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =  -0.3052 (log hazard-ratio)
    hratio =    0.7370
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =      270
    N =      600
    N per group =    300
```

The effect size  $\delta$  now contains the value of the log hazard-ratio.

Continuing with [example 4](#), instead of the estimate of the survival probability in the experimental group, we may have an estimate of the hazard ratio  $\Delta_a$ . For example, the estimate of the hazard ratio in this example is 0.737. We can specify the value of the hazard ratio in the `hratio()` option instead of specifying the experimental-group survival probability 0.6.

```
. power logrank 0.5, onesided hratio(0.737)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    hratio =    0.7370
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =      270
    N =      600
    N per group =    300
```

Alternatively, instead of the hazard ratio we can specify the log hazard-ratio in option `lnhratio()`.

```
. power logrank 0.5, onesided lnhratio(-0.3052)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    ln(hratio) = -0.3052
Censoring:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated number of events and sample sizes:
    E =    270
    N =    600
    N per group =    300
```

The results are identical to the prior results. The estimate of the log hazard-ratio is now displayed in the output instead of the hazard ratio.



## Withdrawal of subjects from the study

Under administrative censoring, the subject is known to have experienced either of the two outcomes by the end of the study: survival or failure. Often, in practice, subjects may withdraw from the study before it terminates and therefore may not experience an event by the end of the study (or be censored) for nonadministrative reasons. Withdrawal of subjects from a study may greatly affect the estimate of the sample size and must be accounted for in the computations. Refer to [Survival data in \[PSS-2\] Intro \(power\)](#) and [\[PSS-5\] Glossary](#) for a formal definition of [withdrawal](#).

[Freedman \(1982\)](#) suggests a conservative adjustment for the estimate of the sample size in the presence of withdrawal, which is implemented in `power logrank`. Withdrawal is assumed to be independent of failure (event) times and administrative censoring.

The proportion of subjects anticipated to withdraw from a study may be specified by using `wdprob()`.

### ▷ Example 7: Withdrawal of subjects from the study

Continuing with [example 4](#), suppose that a withdrawal rate of 10% is expected in the study of colon cancer patients. To account for this, we also specify `wdprob(0.1)`.

```
. power logrank 0.5 0.6, onesided wdprob(0.1)
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
    power =    0.8000
    delta =    0.7370 (hazard ratio)
    hratio =    0.7370
Censoring and withdrawal:
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
    Pr_w =    0.1000
Estimated number of events and sample sizes:
    E =    270
    N =    666
    N per group =    333
```

The estimate of the total sample size using the Freedman method increases from 600 to 666 when the withdrawal rate is assumed to be 10%. The adjustment of the estimate of the sample size for the withdrawal of subjects is conservative. It assumes equal withdrawals from each group; that is, 10% of subjects are lost by the end of the study in each group. This adjustment affects only the estimates of the sample sizes but not the number of events. This is because withdrawal is assumed to be independent of event times, and the ratio of subjects surviving until the end of the study in the two groups does not change under equal withdrawals.

◀

## Including information about subject accrual

Many clinical studies have an accrual period of  $r$ , during which the subjects are recruited to the study, and a follow-up period of  $f = T - r$ , during which the subjects are followed until the end of the study,  $T$ , and no new subjects enter the study. The information about the duration of the accrual and follow-up periods affects the probability of a subject experiencing an event or failing during the study.

[Freedman \(1982\)](#) suggests approximating the combined event-free probability as an average of the survival probabilities in the control and the experimental groups at the minimum follow-up time,  $t^* = f$  (the default approach used in `power logrank`). However, for a long accrual of subjects, this approach may overestimate the required number of subjects, often seriously, because it does not take into account the information about subject follow-up beyond time  $f$ . Here [Freedman \(1982\)](#) proposes to use the survival probabilities at the average follow-up time, defined as  $t^* = (f + T)/2 = f + 0.5r$ , instead of the minimum follow-up time,  $f$ .

Alternatively, [Schoenfeld \(1983\)](#) (see also [Collett \[2015, 479\]](#)) presents a formula for the required number of subjects allowing for uniform accrual (entry, recruitment) over  $[0, r]$  and a follow-up period,  $f$ . This information is incorporated into the formula for the probability of an event (or failure). The formula involves the integrals of the survivor functions of the control and the experimental groups. [Schoenfeld \(1983\)](#) suggests approximating the integral by using Simpson's rule, which requires the



estimates of the survivor function at three specific time points:  $f$ ,  $0.5r + f$ , and  $T = r + f$ . It is sufficient to provide the estimates of these three survival probabilities,  $S_1(f)$ ,  $S_1(0.5r + f)$ , and  $S_1(T)$ , for the control group only. The corresponding survival probabilities of the experimental group are automatically computed using the value of the hazard ratio in `hratio()` (or `log hazard-ratio` in `lnhratio()`) and the proportional-hazards assumption.

The three estimates of the survival probabilities of the control group may be supplied by using the `simpson()` option to adjust the estimates of the sample size or power for uniform entry and a follow-up period. If the estimate of the survivor function over an array of values in the range  $[f, T]$  is available from a previous study, it can be supplied using the `st1()` option to form a more accurate approximation of the probability of an event using numerical integration (see [R] [dydx](#)). Here the value of the length of the accrual period is needed for the computation. It is computed as the difference between the maximum and the minimum values of the time variable  $varname_t$ , supplied using `st1()`, that is,  $r = T - f = \max(varname_t) - \min(varname_t)$ .

For more information, see [Cleves, Gould, and Marchenko \(2016, sec. 16.2\)](#).

### ► Example 8: Sample size in the presence of accrual and follow-up periods

Consider an example described in [Collett \(2015, 482\)](#) of a survival study of chronic active hepatitis. A new treatment is to be compared with a standard treatment with respect to the survival times of the patients with this disease. The investigators want to detect a change in a hazard ratio of 0.57 with 90% power and a 5% two-sided significance level. Also subjects are to be entered into the study uniformly over a period of 18 months and then followed for 24 months. From the Kaplan–Meier estimate of the survivor function available for the control group, the survival probabilities at  $f = 24$ ,  $0.5r + f = 33$ , and  $T = 42$  months are 0.70, 0.57, and 0.45, respectively.

```
. power logrank, hratio(0.57) power(0.9) schoenfeld simpson(0.7 0.57 0.45)
note: probability of an event is computed using Simpson's rule with
      S1(f) = 0.70, S1(f+r/2) = 0.57, S1(T) = 0.45
      S2(f) = 0.82, S2(f+r/2) = 0.73, S2(T) = 0.63
```

```
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
HO: ln(HR) = 0 versus Ha: ln(HR) != 0
```

Study parameters:

```
alpha = 0.0500
power = 0.9000
delta = -0.5621 (log hazard-ratio)
hratio = 0.5700
```

Censoring:

```
Pr_E = 0.3514
```

Estimated number of events and sample sizes:

```
E = 134
N = 380
N per group = 190
```

[Collett \(2015, 309\)](#) reports the required number of events to be 133, which, apart from rounding, agrees with our estimate of 134. In a later example, [Collett \(2003, 309\)](#) uses the number of events, rounded to 140, to compute the required sample size as  $140/0.35 = 400$ , where 0.35 is the estimate of the combined probability of an event. By hand, without rounding the number of events, we compute the required sample size as  $133/0.35 = 380$  and obtain the same estimate of the total sample size as in the output.

Using the average follow-up time suggested by [Freedman \(1982\)](#), we obtain the following:

```
. power logrank 0.57, hratio(0.57) power(0.9) schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
HO: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =  -0.5621 (log hazard-ratio)
    hratio =    0.5700
Censoring:
    s1 =    0.5700
    s2 =    0.7259
    Pr_E =    0.3521
Estimated number of events and sample sizes:
    E =    134
    N =    378
    N per group =    189
```

We specify the survival probability in the control group at  $t^* = 0.5r + f = 0.5 \times 18 + 24 = 33$  as  $S_1(33) = 0.57$  and the hazard ratio of 0.57 (coincidentally). The survival probability in the experimental group is  $S_2(33) = S_1(33)^\Delta = 0.57^{0.57} = 0.726$ . Here we obtain the estimate of the sample size, 378, which is close to the estimate of 380 computed using the more complicated approximation. In this example, the two approximations produce similar results, but this may not always be the case.

The approximation suggested by [Schoenfeld \(1983\)](#) and [Collett \(2015\)](#) is considered to be more accurate because it takes into account information about the patient survival beyond the average follow-up time. In general, the [Freedman \(1982\)](#) and [Schoenfeld \(1983\)](#) approximations tend to give similar results when  $\{\tilde{S}(f) + \tilde{S}(T)\}/2 \approx \tilde{S}(0.5r + f)$ ; see [Methods and formulas](#) for a formal definition of  $\tilde{S}(\cdot)$ .

If we use the survival probability in the control group,  $S_1(24) = 0.7$ , at a follow-up time  $t^* = f = 24$  instead of the average follow-up time  $t^* = 33$  in the presence of an accrual period,

```
. power logrank 0.7, hratio(0.57) power(0.9) schoenfeld
Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
HO: ln(HR) = 0 versus Ha: ln(HR) != 0
Study parameters:
    alpha =    0.0500
    power =    0.9000
    delta =  -0.5621 (log hazard-ratio)
    hratio =    0.5700
Censoring:
    s1 =    0.7000
    s2 =    0.8160
    Pr_E =    0.2420
Estimated number of events and sample sizes:
    E =    134
    N =    550
    N per group =    275
```

we obtain the estimate of the total sample size of 550, which is substantially greater than the previously estimated sample sizes of 380 and 378.

## Computing power

Sometimes the number of subjects available for the enrollment into the study is limited. In such cases, the researchers may want to investigate with what power they can detect a desired treatment effect for a given sample size.

To compute power, you must specify the sample size in the `n()` option and an effect size (a hazard ratio or a log hazard-ratio). A hazard ratio of 0.5 is assumed if an effect size is not specified.

### ► Example 9: Power determination

Recall the colon cancer study described in [example 4](#). Suppose that only 100 subjects are available to be recruited to the study. We find out how this affects the power to detect a hazard ratio of 0.737.

```
. power logrank 0.5, hratio(0.737) onesided n(100)
Estimated power for two-sample comparison of survivor functions
Log-rank test, Freedman method
HO: HR = 1 versus Ha: HR < 1
Study parameters:
    alpha =    0.0500
      N =     100
N per group =    50
    delta =    0.7370 (hazard ratio)
    hratio =    0.7370
Number of events and censoring:
      E =      46
    s1 =    0.5000
    s2 =    0.6000
    Pr_E =    0.4500
Estimated power:
    power =    0.2646
```

The power to detect an alternative  $H_a: \Delta = 0.737$  decreased from 0.8 to 0.2646 when the sample size decreased from 600 to 100 (the number of events decreased from 270 to 46).

◀

### ► Example 10: Multiple values of study parameters

Continuing with [example 9](#), suppose we want to consider a range of sample sizes. We can specify a list (see [\[U\] 11.1.8 numlist](#)) of sample sizes in the `n()` option. For simplicity, we display only power, sample size, and number of events in the table.

```
. power logrank 0.5, hratio(0.737) onesided n(100(100)600) table(power N E)
Estimated power for two-sample comparison of survivor functions
Log-rank test, Freedman method
HO: HR = 1 versus Ha: HR < 1
```

| power | N   | E   |
|-------|-----|-----|
| .2646 | 100 | 46  |
| .4174 | 200 | 91  |
| .5455 | 300 | 136 |
| .6505 | 400 | 181 |
| .7344 | 500 | 226 |
| .8004 | 600 | 271 |

As the sample size increases, the power increases. The decrease in sample size reduces the number of events observed in the study and therefore changes the estimates of the power. If the number of events were fixed, power would have been independent of the sample size, provided that all other parameters were held constant, because the formulas relate power directly to the number of events and not the number of subjects.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS-2] **power, table**. If you wish to produce a power plot, see [PSS-2] **power, graph**.

◀

## Computing effect size

Effect size  $\delta$  for the log-rank test comparing two survivor functions is defined as a hazard ratio (or a log hazard-ratio) of the experimental group to the control group. This value typically defines the clinically significant improvement of the experimental procedure over the control procedure desired to be detected by the log-rank test with a certain power.

Sometimes, we may be interested in determining the smallest effect that yields a statistically significant result for prespecified sample size and power. In this case, both power and sample size must be specified in options `power()` and `n()`, respectively. Additionally, you may also choose the direction of the effect by specifying the `direction()` option. `direction(lower)` is the default, and it assumes  $\Delta_a < 1$  [or  $\ln(\Delta_a) < 0$ ]. You can use `direction(upper)` to compute  $\Delta_a > 1$  [or  $\ln(\Delta_a) > 0$ ].

### ► Example 11: Effect-size determination

Continuing with [example 10](#), we can find that the value of the hazard ratio that can be detected for a fixed sample size of 100 with 80% power is approximately 0.42, corresponding to an increase in survival probability from 0.5 to roughly 0.75.

```
. power logrank 0.5, onesided n(100) power(0.8)
Performing iteration ...
Estimated hazard ratio for two-sample comparison of survivor functions
Log-rank test, Freedman method
H0: HR = 1 versus Ha: HR < 1
Study parameters:
      alpha =    0.0500
      power =    0.8000
         N =      100
  N per group =    50
Number of events and censoring:
      E =      38
      s1 =    0.5000
      s2 =    0.7455
      Pr_E =    0.3772
Estimated effect size and hazard ratio:
      delta =    0.4237 (hazard ratio)
      hratio =    0.4237
```

Under the censoring information, `power logrank` also reports the experimental-group survival rate at the end of the study corresponding to the computed hazard ratio—`s2=0.7455` in our example.

◀

## Testing a hypothesis about two survivor functions using the log-rank test

### ► Example 12: Using the log-rank test to detect a change in survival in two groups

Similarly to [example 4](#), consider the generated dataset `drug.dta`, consisting of variables `drug` (a drug type) and `failtime` (a time to failure).

```
. use https://www.stata-press.com/data/r18/drug
(Patient survival in drug trial)
```

```
. tabulate drug
```

| Treatment type | Freq. | Percent | Cum.   |
|----------------|-------|---------|--------|
| Placebo        | 50    | 33.33   | 33.33  |
| Drug A         | 50    | 33.33   | 66.67  |
| Drug B         | 50    | 33.33   | 100.00 |
| Total          | 150   | 100.00  |        |

```
. by drug, sort: summarize failtime
```

```
-> drug = Placebo
```

| Variable | Obs | Mean    | Std. dev. | Min      | Max      |
|----------|-----|---------|-----------|----------|----------|
| failtime | 50  | 1.03876 | .5535538  | .1687701 | 2.382302 |

```
-> drug = Drug A
```

| Variable | Obs | Mean     | Std. dev. | Min      | Max      |
|----------|-----|----------|-----------|----------|----------|
| failtime | 50  | 1.191802 | .5927507  | .2366922 | 2.277536 |

```
-> drug = Drug B
```

| Variable | Obs | Mean     | Std. dev. | Min      | Max      |
|----------|-----|----------|-----------|----------|----------|
| failtime | 50  | 1.717314 | .8350659  | .5511715 | 3.796102 |

Failure times of the control group (Placebo) were generated from the Weibull distribution with  $\lambda_w = 0.693$  and  $p = 2$  (see [\[ST\] streg](#)); failure times of the two experimental groups, Drug A and Drug B, were generated from Weibull distributions with hazard functions proportional to the hazard of the control group in ratios 0.737 and 0.42, respectively. The Weibull family of survival distributions is chosen arbitrarily, and the Weibull parameter,  $\lambda_w$ , is chosen such that the survival at 1 year,  $t = 1$ , is roughly equal to 0.5. Subjects are randomly allocated to one of the three groups in equal proportions. Subjects with failure times greater than  $t = 1$  will be censored at  $t = 1$ .

Before analyzing these survival data, we need to set up the data using `stset`. After that, we can use `sts test`, `logrank` to test the survivor functions separately for Drug A against Placebo and Drug B against Placebo by using the log-rank test. See [\[ST\] stset](#) and [\[ST\] sts test](#) for more information about these two commands.

```

. stset failtime, exit(time 1)
Survival-time data settings
    Failure event: (assumed to fail at time=failtime)
Observed time interval: (0, failtime]
    Exit on or before: time 1

```

---

```

150 total observations
  0 exclusions

```

---

```

150 observations remaining, representing
 59 failures in single-record/single-failure data
128.985 total analysis time at risk and under observation
                At risk from t =      0
                Earliest observed entry t =      0
                Last observed exit t =      1

```

```

. sts test drug if drug!=2, logrank
    Failure _d: 1 (meaning all fail)
    Analysis time _t: failtime
    Exit on or before: time 1
Equality of survivor functions
Log-rank test

```

| drug    | Observed events | Expected events |
|---------|-----------------|-----------------|
| Placebo | 25              | 22.17           |
| Drug A  | 21              | 23.83           |
| Total   | 46              | 46.00           |

```

                chi2(1) = 0.70
                Pr>chi2 = 0.4028

```

```

. sts test drug if drug!=1, logrank
    Failure _d: 1 (meaning all fail)
    Analysis time _t: failtime
    Exit on or before: time 1
Equality of survivor functions
Log-rank test

```

| drug    | Observed events | Expected events |
|---------|-----------------|-----------------|
| Placebo | 25              | 16.61           |
| Drug B  | 13              | 21.39           |
| Total   | 38              | 38.00           |

```

                chi2(1) = 7.55
                Pr>chi2 = 0.0060

```

From the results from `sts test` for the Drug A group, we fail to reject the null hypothesis of no difference between the survivor functions in the two groups; given our simulated data, the test made a type II error. On the other hand, for the Drug B group the one-sided  $p$ -value of 0.003, computed as  $0.006/2 = 0.003$ , suggests that the null hypothesis of nonsuperiority of the experimental treatment be rejected at the 0.005 significance level. We correctly conclude that the data provide the evidence that Drug B is superior to the Placebo.

Results from `sts test`, `logrank` for the two experimental groups agree with findings from examples 9 and 11. For the sample size of 100, the power of the log-rank test to detect the hazard ratio of 0.737 (10% increase in survival) is low (26%), whereas this sample size is sufficient for the test to detect a change in a hazard of 0.42 (25% increase in survival) with approximately 80% power.

Here we simulated our data from the alternative hypothesis and therefore can determine whether the correct decision or a type II error was made by the test. In practice, however, there is no way to determine the accuracy of the decision from the test. All we know is that in a long series of trials, there is a 5% chance that a particular test will incorrectly reject the null hypothesis, a 74%  $[(1 - 0.25) \times 100\%]$  given the power of 0.2646 obtained in [example 9](#)] chance that the test will miss the alternative  $H_a$ :  $\Delta = 0.737$ , and a 20%  $[(1 - 0.8) \times 100\%]$  given the power of 0.8 in [example 11](#)] chance that the test will miss the alternative  $H_a$ :  $\Delta = 0.42$ .

◀

## Stored results

power logrank stores the following in `r()`:

### Scalars

|                             |  |
|-----------------------------|--|
| <code>r(alpha)</code>       | significance level   |
| <code>r(power)</code>       | power  |
| <code>r(beta)</code>        | probability of a type II error                                   |
| <code>r(delta)</code>       | effect size  |
| <code>r(N)</code>           | total sample size  |
| <code>r(N_a)</code>         | actual sample size   |
| <code>r(N1)</code>          | sample size of the control group                                 |
| <code>r(N2)</code>          | sample size of the experimental group                            |
| <code>r(nratio)</code>      | ratio of sample sizes, $N2/N1$                                   |
| <code>r(nratio_a)</code>    | actual ratio of sample sizes                                     |
| <code>r(nfractional)</code> | 1 if <code>nfractional</code> is specified, 0 otherwise          |
| <code>r(onesided)</code>    | 1 for a one-sided test, 0 otherwise                              |
| <code>r(E)</code>           | total number of events (failures)                                |
| <code>r(hratio)</code>      | hazard ratio   |
| <code>r(lnhratio)</code>    | log hazard-ratio   |
| <code>r(s1)</code>          | survival probability in the control group (if specified)         |
| <code>r(s2)</code>          | survival probability in the experimental group (if specified)    |
| <code>r(Pr_E)</code>        | probability of an event (failure)                                |
| <code>r(Pr_w)</code>        | proportion of withdrawals  |
| <code>r(t_min)</code>       | minimum time (if <code>st1()</code> is specified)                |
| <code>r(t_max)</code>       | maximum time (if <code>st1()</code> is specified)                |
| <code>r(separator)</code>   | number of lines between separator lines in the table             |
| <code>r(divider)</code>     | 1 if <code>divider</code> is requested in the table, 0 otherwise |
| <code>r(init)</code>        | initial value for hazard ratio or log hazard-ratio               |
| <code>r(maxiter)</code>     | maximum number of iterations                                     |
| <code>r(iter)</code>        | number of iterations performed                                   |
| <code>r(tolerance)</code>   | requested parameter tolerance                                    |
| <code>r(deltax)</code>      | final parameter tolerance achieved                               |
| <code>r(ftolerance)</code>  | requested distance of the objective function from zero           |
| <code>r(function)</code>    | final distance of the objective function from zero               |
| <code>r(converged)</code>   | 1 if iteration algorithm converged, 0 otherwise                  |

### Macros

|                           |   |
|---------------------------|---|
| <code>r(type)</code>      | test  |
| <code>r(method)</code>    | logrank   |
| <code>r(test)</code>      | Freedman or Schoenfeld  |
| <code>r(effect)</code>    | <code>hratio</code> or <code>lnhratio</code>  |
| <code>r(survvar)</code>   | name of the variable containing survival probabilities (if <code>st1()</code> is specified) |
| <code>r(timevar)</code>   | name of the variable containing time points (if <code>st1()</code> is specified)            |
| <code>r(direction)</code> | lower or upper  |
| <code>r(columns)</code>   | displayed table columns   |
| <code>r(labels)</code>    | table column labels   |
| <code>r(widths)</code>    | table column widths   |
| <code>r(formats)</code>   | table column formats  |

### Matrices

|                           |   |
|---------------------------|---|
| <code>r(pss_table)</code> | table of results  |
| <code>r(simpmat)</code>   | control-group survival probabilities (if <code>simpson()</code> is specified) |

## Methods and formulas

Let  $S_1(t)$  and  $S_2(t)$  denote the survivor functions of the control and the experimental groups and  $\Delta(t) = \ln\{S_2(t)\}/\ln\{S_1(t)\}$  denote the hazard ratio at time  $t$  of the experimental to the control groups. Thus, for a given constant hazard ratio  $\Delta$ , the survivor function of the experimental group at any time  $t > 0$  may be computed as  $S_2(t) = \{S_1(t)\}^\Delta$  under the assumption of proportional hazards. Define  $E$  and  $n$  to be the total number of events and the total number of subjects required for the study, respectively;  $p_w$  to be the proportion of subjects withdrawn from the study (lost to follow-up); and  $z_{(1-\alpha/k)}$  and  $z_{(1-\beta)}$  to be the  $(1 - \alpha/k)$ th and the  $(1 - \beta)$ th quantiles of the standard normal distribution, respectively, with  $k = 1$  for the one-sided test and  $k = 2$  for the two-sided test. Let  $R$  be the allocation ratio to the experimental group with respect to the control group, that is,  $n_2 = Rn_1$ .

The total number of events required to be observed in a study to ensure a power of  $\pi = 1 - \beta$  of the log-rank test to detect the hazard ratio  $\Delta$  with significance level  $\alpha$ , according to [Freedman \(1982\)](#), is

$$E = \frac{1}{R}(z_{1-\alpha/k} + z_{1-\beta})^2 \left( \frac{R\Delta + 1}{\Delta - 1} \right)^2$$

and, according to [Schoenfeld \(1983\)](#) and [Collett \(2015, 473\)](#), is

$$E = \frac{1}{R}(z_{1-\alpha/k} + z_{1-\beta})^2 \left\{ \frac{1+R}{\ln(\Delta)} \right\}^2$$

Both formulas are approximations and rely on a set of assumptions such as distinct failure times, all subjects completing the course of the study (no withdrawal), and a constant ratio,  $R$ , of subjects at risk in two groups at each failure time.

The total sample size required to observe the total number of events,  $E$ , is given by

$$n = \frac{E}{p_E}$$

The number of subjects required to be recruited in each group is obtained as  $n_1 = n/(1 + R)$  and  $n_2 = nR/(1 + R)$ . If `nfractional` is not specified, sample sizes and the number of events are rounded to integer values; see [Fractional sample sizes](#) in [\[PSS-4\] Unbalanced designs](#) for details.

By default, the probability of an event (failure),  $p_E$ , is approximated as suggested by [Freedman \(1982\)](#),

$$p_E = 1 - \frac{S_1(t^*) + RS_2(t^*)}{1 + R}$$

where  $t^*$  is the minimum follow-up time,  $f$ , or, in the presence of an accrual period, the average follow-up time,  $(f + T)/2 = f + 0.5r$ .

If `simpson()` is specified, the probability of an event is approximated using Simpson's rule as suggested by [Schoenfeld \(1983\)](#):

$$p_E = 1 - \frac{1}{6} \left\{ \tilde{S}(f) + 4\tilde{S}(0.5r + f) + \tilde{S}(T) \right\}$$

where  $\tilde{S}(t) = \{S_1(t) + RS_2(t)\}/(1 + R)$  and  $f$ ,  $r$ , and  $T = f + r$  are the follow-up period, the accrual period, and the total duration of the study, respectively.



The methods do not incorporate time explicitly but rather use it to determine values of the survival probabilities  $S_1(t)$  and  $S_2(t)$  used in the computations.

If `st1()` is used, the integral in the expression for the probability of an event

$$p_E = 1 - \frac{1}{r} \int_f^T \tilde{S}(t) dt$$

is computed numerically using cubic splines (see [R] `dydx`). The value of  $r$  is computed as the difference between the maximum and the minimum values of `varnamet` in `st1()`,  $r = T - f = \max(\text{varname}_t) - \min(\text{varname}_t)$ .

To account for the proportion of subjects,  $p_w$ , withdrawn from the study (lost to follow-up), a conservative adjustment to the total sample size is applied as follows:

$$n_w = \frac{n}{1 - p_w}$$

Equal withdrawal rates are assumed in the adjustment of the group sample sizes for the withdrawal of subjects. Equal withdrawals do not affect the estimates of the number of events, provided that withdrawal is independent of event times and the ratio of subjects at risk in two groups remains constant at each failure time.

The power for each method is estimated using the formula

$$\pi = 1 - \beta = \Phi\{|\psi|^{-1}(Rnp_E)^{1/2} - z_{1-\alpha/k}\}$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function;  $\psi = (R\Delta + 1)/(\Delta - 1)$  or  $\psi = (1 + R)/\ln(\Delta)$  if the `schoenfeld` option is specified.

The estimate of the hazard ratio (or log hazard-ratio) for fixed power and sample size is computed (iteratively for censoring) using the formulas for the sample size given above. The value of the hazard ratio (log hazard-ratio) corresponding to the reduction in a hazard of the experimental group relative to the control group is reported by default.

## References

- Cleves, M. A., W. W. Gould, and Y. V. Marchenko. 2016. *An Introduction to Survival Analysis Using Stata*. Rev. 3rd ed. College Station, TX: Stata Press.
- Collett, D. 2015. *Modelling Survival Data in Medical Research*. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC.
- Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129. <https://doi.org/10.1002/sim.4780010204>.
- Hsieh, F. Y. 1992. Comparing sample size formulae for trials with unbalanced allocation using the logrank test. *Statistics in Medicine* 11: 1091–1098. <https://doi.org/10.1002/sim.4780110810>.
- Lakatos, E., and K. K. G. Lan. 1992. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine* 11: 179–191. <https://doi.org/10.1002/sim.4780110205>.
- Machin, D., M. J. Campbell, S. B. Tan, and S. H. Tan. 2009. *Sample Size Tables for Clinical Studies*. 3rd ed. Chichester, UK: Wiley-Blackwell.
- Marubini, E., and M. G. Valsecchi. 1995. *Analysing Survival Data from Clinical Trials and Observational Studies*. New York: Wiley.
- Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68: 316–319. <https://doi.org/10.2307/2335833>.

—. 1983. Sample-size formula for the proportional-hazards regression model. *Biometrics* 39: 499–503. <https://doi.org/10.2307/2531021>.

Schoenfeld, D. A., and J. R. Richter. 1982. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* 38: 163–170. <https://doi.org/10.2307/2530299>.

Also see [PSS-2] **Intro (power)** for more references.

## Also see

[PSS-2] **power logrank, cluster** — Power analysis for the log-rank test, CRD

[PSS-2] **power** — Power and sample-size analysis for hypothesis tests

[PSS-2] **power cox** — Power analysis for the Cox proportional hazards model

[PSS-2] **power exponential** — Power analysis for a two-sample exponential test

[PSS-2] **power, graph** — Graph results from the power command

[PSS-2] **power, table** — Produce table of results from the power command

[PSS-5] **Glossary**

[ADAPT] **gsdesign logrank** — Group sequential design for a log-rank test

[ST] **stcox** — Cox proportional hazards model

[ST] **sts test** — Test equality of survivor functions

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

