Estimating risks with the new strisk command

Matteo Bottai, ScD

Karolinska Institutet
Stockholm, Sweden

Patient survival in drug trial

We use the Cancer dataset.

```
. sysuse cancer, clear
(Patient Survival in Drug Trial)
. describe studytime died drug
```

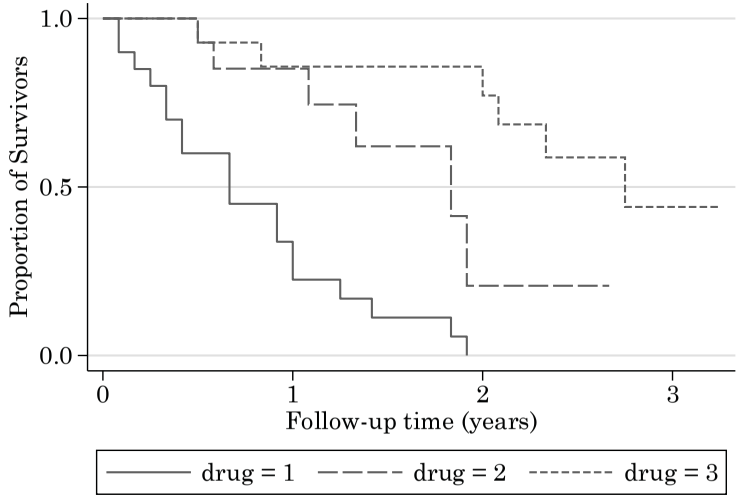
variable name	storage type	display format	value label	variable label
studytime	byte	%8.0g		Months to death or end of exp.
died	byte	%8.0g		1 if patient died
drug	byte	%8.0g		Drug type (1=placebo)

Patient survival in drug trial

We define the time and failure variables.
We reset the time scale to years.

```
. stset studytime, failure(died) scale(12)
     failure event:  died != 0 & died < .
obs. time interval:  (0, studytime]
 exit on or before:  failure
    t for analysis:  time/12
--------------------------------------------------------------------------
        48  total observations
         0  exclusions
--------------------------------------------------------------------------
        48  observations remaining, representing
        31  failures in single-record/single-failure data
        62  total analysis time at risk and under observation
                                        at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      3.25
```


Patient survival in drug trial



Mortality rates

We calculate mortality *rates* with the strate command.

```
. strate drug
        failure _d:  died
  analysis time _t:  studytime/12
Estimated rates and lower/upper bounds of 95% confidence intervals
(48 records included in the analysis)
```

drug	D	Y	Rate	Lower	Upper
1	19	15.0000	1.266667	0.807948	1.985827
2	6	17.4167	0.344498	0.154769	0.766810
3	6	29.5833	0.202817	0.091118	0.451446

The incidence rate in group 1 is 1.27 (95% CI 0.81, 1.99).
A patient in group 1 is expected to die 1.27 times per year.

Mortality risks

We calculate mortality *risks* the new strisk command.

```
. strisk drug
   Incidence risk
```

drug	Risk	[95% Conf.	Int.]
1	0.830235	0.578028	0.973872
2	0.381161	0.163559	0.724608
3	0.206644	0.090824	0.430367

The incidence risk in group 1 is 0.83.
A patient in group 1 has a yearly mortality risk of 0.83.
The risks and confidence intervals are between zero and one.

Stata syntax

The syntax of strisk is similar to that of strate.

`strisk` [*varlist*] [*if*] [*in*] [, `level(#) graph nowhisker`]

strisk tabulates risks by the variables in *varlist*, if specified.
`level` specifies the level of the confidence intervals.
`graph` plots the risks by *varlist*, if specified.
`nowhisker` omits the confidence intervals from the graph.

Incidence risks

The probability of occurrence of an event over the time interval (t_0, t_1) conditional on $T > t_0$ is

$$\Pr[T \leq t_1 \mid T > t_0] = 1 - \frac{S(t_1)}{S(t_0)}$$

where $S(t) = P(T > t)$ is the survival function.

Bottai (2017) defined the average probability of occurrence of an event over the interval (t_0, t_1) conditional on $T > t_0$ as

$$G(t_0, t_1) = 1 - \left[\frac{S(t_1)}{S(t_0)}\right]^{1/(t_1 - t_0)}$$

An example

The event of interest is death between $t_0 = 0$ and $t_1 = 3$.
The risk of dying is the complement of the risk of surviving,

$$P(T \leq t_1 | T > t_0) = 1 - P(T > t_1 | T > t_0) = 1 - \frac{S(3)}{S(0)}$$

The average probability of surviving each unit time is

$$\left[\frac{S(1)}{S(0)}\ \frac{S(2)}{S(1)}\ \frac{S(3)}{S(2)}\right]^{1/3} = \left[\frac{S(3)}{S(0)}\right]^{1/3}$$

The average probability of dying per unit time is

$$G(0,3) = 1 - \left[\frac{S(3)}{S(0)}\right]^{1/3}$$

The strisk command

The risk can be written as

$$G(t_0, t_1) = 1 - \exp\left[H(t_0) - H(t_1)\right] / (t_1 - t_0)$$

where $H(t) = -\log S(t)$ is the cumulative hazard function.

The strisk command uses the above expression.
It sets t_0 to the start of the follow-up time.
It sets t_1 to the largest observed time value.
It sets $H(t_0)$ to zero, assuming everyone is alive at start.
It calculates $H(t_1)$ with sts list with the Nelson-Aalen option.

Final comments

The new strisk command

– calculates risks of events with survival data
– is consistent with the st suite of commands
– is computationally fast
– should be preferred to strate in evaluating risks

To download the command and help files, type

```
net install strisk, from(http://www.imm.ki.se/biostatistics/stata)
```