

## Title

**intro** — Introduction to survival analysis manual

## Description

This entry describes this manual and what has changed since Stata 10. See the next entry, [ST] **survival analysis**, for an introduction to Stata’s survival analysis capabilities.

## Remarks

This manual documents commands for survival analysis and epidemiological tables and is referred to as [ST] in cross-references. Following this entry, [ST] **survival analysis** provides an overview of the commands.

This manual is arranged alphabetically. If you are new to Stata’s survival analysis and epidemiological tables commands, we recommend that you read the following sections first:

[ST] <b>survival analysis</b>	Introduction to survival analysis & epidemiological tables commands
[ST] <b>st</b>	Survival-time data
[ST] <b>stset</b>	Set variables for survival data

Stata is continually being updated, and Stata users are always writing new commands. To find out about the latest survival analysis features, type `search survival` after installing the latest official updates; see [R] **update**. To find out about the latest epidemiological features, type `search epi`.

## What’s new

This section is intended for previous Stata users. If you are new to Stata, you may as well skip it.

1. New command `stcrreg` fits competing-risks regression models for survival data, according to the method of Fine and Gray (1999). In a competing risks model, subjects are at risk of failure because of two or more separate and possibly correlated causes. See [ST] **stcrreg**. Existing command `stcurve` will now graph cumulative incidence functions after `stcrreg`; see [ST] **stcurve**.
2. Stata’s new multiple-imputation commands for dealing with missing values may be used with `stcox`, `streg`, and `stcrreg`; see [MI] **intro**. Either `stset` your data before using `mi set`, or use `mi stset` to `stset` your data afterward.
3. Factor variables may now be used with `stcox`, `streg`, and `stcrreg`. See [U] **11.4.3 Factor variables**.
4. New reporting options `baselevels` and `allbaselevels` control how base levels of factor variables are displayed in output tables. New reporting option `noemptycells` controls whether missing cells in interactions are displayed.

These new options are supported by estimation commands `stcox`, `streg`, and `stcrreg`, and by existing postestimation commands `estat summarize` and `estat vce`. See [R] **estimation options**.

5. New reporting option `noomitted` controls whether covariates that are dropped because of collinearity are reported in output tables. By default, Stata now includes a line in estimation and related output tables for collinear covariates and marks those covariates as “(omitted)”. `noomitted` suppresses those lines.

noomitted is supported by estimation commands `stcox`, `streg`, and `stcrreg`, and by existing postestimation commands `estat summarize` and `estat vce`. See [R] **estimation options**.

6. New option `vsquish` eliminates blank lines in estimation and related tables. Many output tables now set off factor variables and time-series-operated variables with a blank line. `vsquish` removes these lines.

`vsquish` is supported by estimation commands `stcox`, `streg`, and `stcrreg`, and by existing postestimation command `estat summarize`. See [R] **estimation options**.

7. Estimation commands `stcox`, `streg`, and `stcrreg` support new option `coeflegend` to display the coefficients' legend rather than the coefficient table. The legend shows how you would type a coefficient in an expression, in a test command, or in a constraint definition. See [R] **estimation options**.
8. Estimation commands `streg` and `stcrreg` support new option `nocnsreport` to suppress reporting constraints; see [R] **estimation options**.
9. Concerning `predict`:

- a. `predict` after `stcox` offers three new diagnostic measures of influence: `DFBETAS`, likelihood displacement values, and `LMAX` statistics. See [ST] **stcox postestimation**.
- b. `predict` after `stcox` can now calculate diagnostic statistics `basesurv()`, `basechazard()`, `basehc()`, `mgale()`, `effects()`, `esr()`, `schoenfeld()`, and `scaledsch()`. Previously, you had to request these statistics when you fit the model by specifying the option with the `stcox` command. Now you obtain them by using `predict` after estimation. The options continue to work with `stcox` directly but are no longer documented. See [ST] **stcox postestimation**.
- c. `predict` after `stcox` and `streg` now produces subject-level residuals by default. Previously, record-level or partial results were produced, although there was an inconsistency. This affects multiple-record data only because there is no difference between subject-level and partial residuals in single-record data. This change affects `predict`'s options `mgale`, `csnell`, `deviance`, and `scores` after `stcox` (and new options `ldisplace`, `lmax`, and `dfbeta`, of course); and it affects `mgale` and `deviance` after `streg`. `predict`, `deviance` was the inconsistency; it always produced subject-level results.

For instance, in previous Stata versions you typed

```
. predict cs, csnell
```

to obtain partial Cox–Snell residuals. One statistic per record was produced. To obtain subject-level residuals, for which there is one per subject and which `predict` stored on each subject's last record, you typed

```
. predict ccs, ccsnell
```

In Stata 11, when you type

```
. predict cs, csnell
```

you obtain the subject-level residual. To obtain the partial, you use the new `partial` option:

```
. predict cs, csnell partial
```

The same applies to all the other residuals. Concerning the inconsistency, partial deviances are now available.

Not affected is `predict`, `scores` after `streg`. Log-likelihood scores in parametric models are mathematically defined at the record level and are meaningful only if evaluated at that level.

Prior behavior is restored under version control. See [ST] **stcox postestimation**, [ST] **streg postestimation**, and [ST] **stcrreg postestimation**.

10. **stcox** now allows up to 100 time-varying covariates as specified in option `tvf()`. The previous limit was 10. See [ST] **stcox**.
11. Existing commands **stcurve** and **estat phtest** no longer require that you specify the appropriate options to **stcox** before using them. The commands automatically generate the statistics they require. See [ST] **stcurve** and [ST] **stcox PH-assumption tests**.
12. Existing **epitab** commands **ir**, **cs**, **cc**, and **mhodds** now treat missing categories of variables in `by()` consistently. By default, missing categories are now excluded from the computation. This may be overridden by specifying `by()`'s new option `missing`. See [ST] **epitab**.
13. Existing command **sts list** has new option `saving()` that creates a dataset containing the results. See [ST] **sts list**.

For a complete list of all the new features in Stata 11, see [U] **1.3 What's new**.

## Reference

Fine, J. P., and R. J. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 94: 496–509.

## Also see

[U] **1.3 What's new**

[R] **intro** — Introduction to base reference manual