

Notes on Survival analysis using Stata

Jae-Woo Kim

Stata

- Stset
- Sts list
- Sts graph
- Sts generate
- Sts test
- Stphplot
- Stcoxkm
- Stcox
- Stphtest
- streg

steps

- Estimating survival functions (unadjusted) and comparing them across strata
- Assessing the PH assumption using graphs
- Running a Cox PH model
- Running a stratified Cox model
- Assessing the PH assumption with a statistical test
- Obtaining Cox adjusted survival curves
- Running an extended Cox model
- Running parametric models
- Running frailty models
- Modeling recurrent events

Data layout

	id	event1	event2	event3	event4	per1	per2	per3	per4	busi
1	1	0	0	1	0	1	2	3	4	1
2	2	2	0	0	1	2	2	3	3	1
3	3	0	0	0	0	0	1	1	1	1
4	4	0	2	0	0	0	3	3	4	2
5	5	0	1	2	1	1	1	2	2	2

	id	_t0	_t	event	per	busi
1	1	0	1	0	1	1
2	1	1	2	0	2	1
3	1	2	3	1	3	1
4	1	3	4	0	4	1
5	2	0	1	2	2	1
6	2	1	2	0	2	1
7	2	2	3	0	3	1
8	2	3	4	1	3	1
9	3	0	1	0	0	1
10	3	1	2	0	1	1
11	3	2	3	0	1	1
12	3	3	4	0	1	1
13	4	0	1	0	0	2
14	4	1	2	2	3	2
15	4	2	3	0	3	2
16	4	3	4	0	4	2
17	5	0	1	0	1	2
18	5	1	2	1	1	2
19	5	2	3	2	2	2
20	5	3	4	1	2	2

Stset survt, failure(event==1) id(id)

	id	survt	event
1	1	3	1
2	2	4	1
3	3	4	0
4	4	2	1
5	5	3	1

id(idvar) specifies the subject-id variable; idvar may be string or numeric. Observations for which idvar contains missing (. or "") are ignored. If id() is not specified, each observation is assumed to represent a different subject.

	id	survt	event	_st	_d	_t	_t0
1	1	3	1	1	1	3	0
2	2	4	1	1	1	4	0
3	3	4	0	1	0	4	0
4	4	2	1	1	1	2	0
5	5	3	1	1	1	3	0

Stsplit v1, at(1); stsplit v2, at(2)

	id	survt	event	_st	_d	_t	_t0	v1
1	1	1	.	1	0	1	0	0
2	1	3	1	1	1	3	1	1
3	2	1	.	1	0	1	0	0
4	2	4	1	1	1	4	1	1
5	3	1	.	1	0	1	0	0
6	3	4	0	1	0	4	1	1
7	4	1	.	1	0	1	0	0
8	4	2	1	1	1	2	1	1
9	5	1	.	1	0	1	0	0
10	5	3	1	1	1	3	1	1

	id	survt	event	_st	_d	_t	_t0	v1	v2
1	1	1	.	1	0	1	0	0	0
2	1	2	.	1	0	2	1	1	0
3	1	3	1	1	1	3	2	1	2
4	2	1	.	1	0	1	0	0	0
5	2	2	.	1	0	2	1	1	0
6	2	4	1	1	1	4	2	1	2
7	3	1	.	1	0	1	0	0	0
8	3	2	.	1	0	2	1	1	0
9	3	4	0	1	0	4	2	1	2
10	4	1	.	1	0	1	0	0	0
11	4	2	1	1	1	2	1	1	0
12	5	1	.	1	0	1	0	0	0
13	5	2	.	1	0	2	1	1	0
14	5	3	1	1	1	3	2	1	2

stsplot v3, at(3)

	id	survt	event	_st	_d	_t	_t0	v1	v2	v3
1	1	1	.	1	0	1	0	0	0	0
2	1	2	.	1	0	2	1	1	0	0
3	1	3	1	1	1	3	2	1	2	0
4	2	1	.	1	0	1	0	0	0	0
5	2	2	.	1	0	2	1	1	0	0
6	2	3	.	1	0	3	2	1	2	0
7	2	4	1	1	1	4	3	1	2	3
8	3	1	.	1	0	1	0	0	0	0
9	3	2	.	1	0	2	1	1	0	0
10	3	3	.	1	0	3	2	1	2	0
11	3	4	0	1	0	4	3	1	2	3
12	4	1	.	1	0	1	0	0	0	0
13	4	2	1	1	1	2	1	1	0	0
14	5	1	.	1	0	1	0	0	0	0
15	5	2	.	1	0	2	1	1	0	0
16	5	3	1	1	1	3	2	1	2	0

Save the file as as comma separated values (.csv);
 insheet using d:\2tvc_another.csv

	A	B	C	D	E	F	G	H	
1	x1	x2	x3	x4	y1	y2	y3	y4	
2	0	0	1	.	1	2	3	4	
3	0	0	0	1	2	2	3	3	
4	0	0	0	0	0	1	1	1	
5	0	1	.	.	0	3	3	4	
6	0	0	1	.	1	1	2	2	

	x1	x2	x3	x4	y1	y2	y3	y4	
1	0	0	1	.	1	2	3	4	
2	0	0	0	1	2	2	3	3	
3	0	0	0	0	0	1	1	1	
4	0	1	.	.	0	3	3	4	
5	0	0	1	.	1	1	2	2	

	id	year	x	y
1	1	1	0	1
2	1	2	0	2
3	1	3	1	3
4	1	4	.	4
5	2	1	0	2
6	2	2	0	2
7	2	3	0	3
8	2	4	1	3
9	3	1	0	0
10	3	2	0	1
11	3	3	0	1
12	3	4	0	1
13	4	1	0	0
14	4	2	1	3
15	4	3	.	3
16	4	4	.	4
17	5	1	0	1
18	5	2	0	1
19	5	3	1	2
20	5	4	.	2

```
. insheet using d:\2tvc_another.csv
(8 vars, 5 obs)
```

```
. edit
```

```
- preserve
```

```
. gen id=_n
```

```
. reshape long x y, i(id) j(year)
```

```
(note: j = 1 2 3 4)
```

```
Data
```

```
wide -> long
```

```
Number of obs.
```

```
5 -> 20
```

```
Number of variables
```

```
9 -> 4
```

```
j variable (4 values)
```

```
-> year
```

```
xij variables:
```

```
x1 x2 ... x4 -> x
```

```
y1 y2 ... y4 -> y
```

	id	year	x	y
1	1	1	0	1
2	1	2	0	2
3	1	3	1	3
4	2	1	0	2
5	2	2	0	2
6	2	3	0	3
7	2	4	1	3
8	3	1	0	0
9	3	2	0	1
10	3	3	0	1
11	3	4	0	1
12	4	1	0	0
13	4	2	1	3
14	5	1	0	1
15	5	2	0	1
16	5	3	1	2

- Drop if $x==.$

This is why I gave censored data . in the excel file at the beginning.

	id	year	event1	per
1	1	1	0	1
2	1	2	0	2
3	1	3	1	3
4	2	1	0	2
5	2	2	0	2
6	2	3	0	3
7	2	4	1	3
8	3	1	0	0
9	3	2	0	1
10	3	3	0	1
11	3	4	0	1
12	4	1	0	0
13	4	2	1	3
14	5	1	0	1
15	5	2	0	1
16	5	3	1	2

* Rename x event1;
rename y per

Merge!

	id	survt	event	_st	_d	_t	_t0	v1	v2	v3
1	1	1	.	1	0	1	0	0	0	0
2	1	2	.	1	0	2	1	1	0	0
3	1	3	1	1	1	3	2	1	2	0
4	2	1	.	1	0	1	0	0	0	0
5	2	2	.	1	0	2	1	1	0	0
6	2	3	.	1	0	3	2	1	2	0
7	2	4	1	1	1	4	3	1	2	3
8	3	1	.	1	0	1	0	0	0	0
9	3	2	.	1	0	2	1	1	0	0
10	3	3	.	1	0	3	2	1	2	0
11	3	4	0	1	0	4	3	1	2	3
12	4	1	.	1	0	1	0	0	0	0
13	4	2	1	1	1	2	1	1	0	0
14	5	1	.	1	0	1	0	0	0	0
15	5	2	.	1	0	2	1	1	0	0
16	5	3	1	1	1	3	2	1	2	0

	id	year	event1	per
1	1	1	0	1
2	1	2	0	2
3	1	3	1	3
4	2	1	0	2
5	2	2	0	2
6	2	3	0	3
7	2	4	1	3
8	3	1	0	0
9	3	2	0	1
10	3	3	0	1
11	3	4	0	1
12	4	1	0	0
13	4	2	1	3
14	5	1	0	1
15	5	2	0	1
16	5	3	1	2

1. Estimating survival functions (unadjusted) and comparing them across strata

addicts.dta

- Patient ID
- SURVT: the time (days) until the patient dropped out of the clinic or was censored
- STATUS: dropped(1); censored(0)
- CLINIC: indicates which methadone treatment clinic the patient attended
- PRISON: a prison record(1) or not(0)
- DOSE: a continuous variable for the patient's maximum methadone dose (mg/day)

```

. stset survt, failure( status==1) id(id)

      id: id
  failure event: status == 1
obs. time interval: (survt[_n-1], survt]
  exit on or before: failure
-----
 238 total obs.
   0 exclusions
-----
 238 obs. remaining, representing
 238 subjects
 150 failures in single failure-per-subject data
95812 total analysis time at risk, at risk from t = 0
      earliest observed entry t = 0
      last observed exit t = 1076

```

- Stset survt, failure(status==1) id(id)
- ID not needed here b/c each obs represents a different patient, but If there were multiple observations and multiple events for a single subject (cluster), stata can provide robust variance estimates appropriate for clustered data.

- New variables added

_t the time-to-event variable

_d the status variable 1 for event

_t0 the beginning time variable. All observations start at time 0 by default

_st indicates which variables are used in the analysis. All observations are used (coded 1) by default

```

.stdes

      failure_d: status == 1
      analysis time_t: survt
      id: id

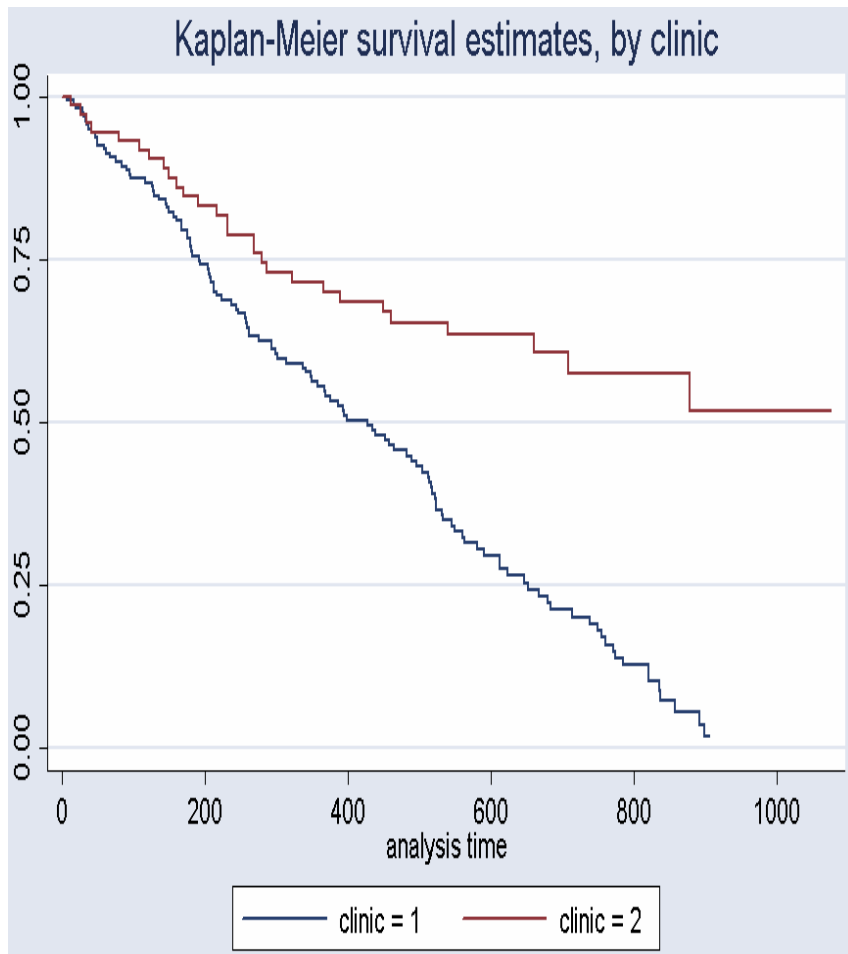
Category          total |----- per subject -----|
                   mean   min   median   max
-----
no. of subjects      238
no. of records       238      1      1      1      1
(first) entry time          0      0      0      0
(final) exit time      402.5714      2    367.5    1076

subjects with gap          0
time on gap if gap          0      .      .      .
time at risk           95812    402.5714      2    367.5    1076

failures              150    .6302521      0      1      1

```

- Stdes
- Sts list (KM survival estimate)



- Sts list, by(clinic) compare at (0 20 to 1080)
- Not (0 to 1080)
- Compare (): specified time points
- Sts graph, by(clinic)
- Sts graph, by(clinic) failure: cumulative risk rather than survival

```
. sts test clinic
```

```
      failure _d: status == 1  
analysis time _t: survt  
              id: id
```

Log-rank test for equality of survivor functions

clinic	Events observed	Events expected
1	122	90.91
2	28	59.09
Total	150	150.00
	chi2(1) =	27.89
	Pr>chi2 =	0.0000

- Sts test clinic: default (log rank test)
- Sts test clinic, wilcoxon
- Sts test clinic, tware
- Sts test clinic, peto
- Sts test clinic, strata(prison)
- Sts generate skm=s, by(clinic): create a new variable in the dataset containing KM survival estimates

```
. ltable survt status, by(clinic)
```

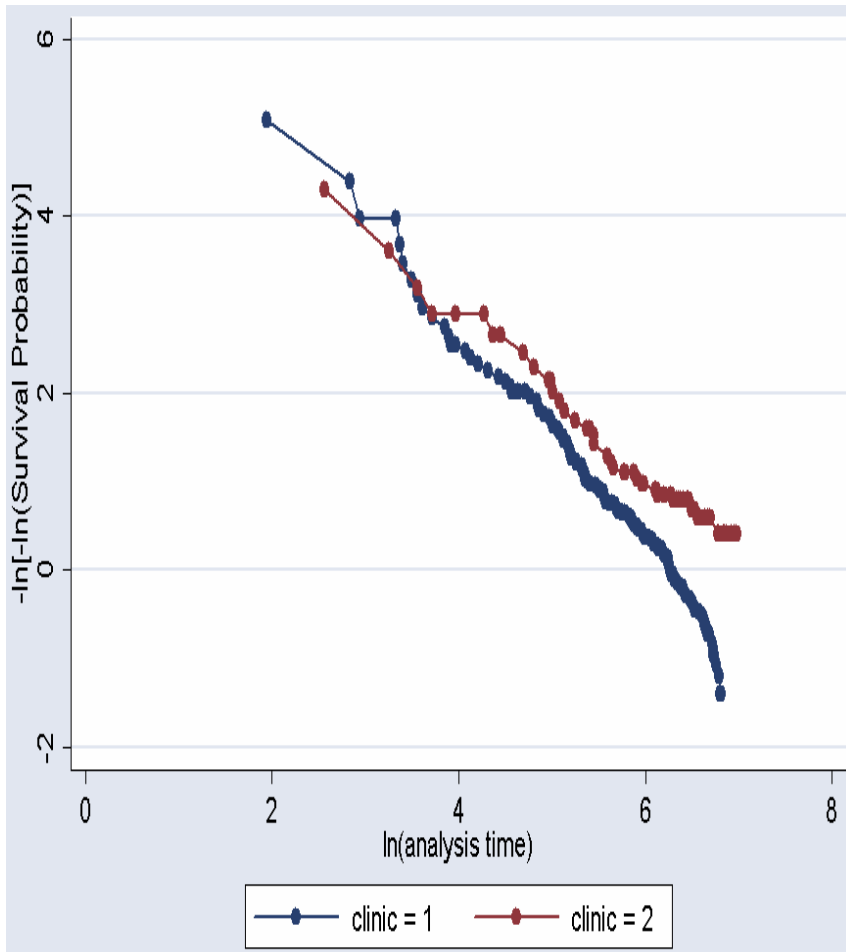
Interval	Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
clinic = 1							
2 3	163	0	1	1.0000	0.0000	.	.
7 8	162	1	0	0.9938	0.0062	0.9570	0.9991
17 18	161	1	0	0.9877	0.0087	0.9515	0.9969
19 20	160	1	0	0.9815	0.0106	0.9437	0.9940
28 29	159	0	2	0.9815	0.0106	0.9437	0.9940
29 30	157	1	0	0.9752	0.0122	0.9354	0.9906
30 31	156	1	0	0.9690	0.0137	0.9271	0.9870
33 34	155	1	0	0.9627	0.0149	0.9189	0.9831
35 36	154	1	0	0.9565	0.0161	0.9109	0.9790
37 38	153	1	0	0.9502	0.0172	0.9029	0.9748
41 42	152	1	0	0.9440	0.0181	0.8951	0.9704
47 48	151	1	0	0.9377	0.0191	0.8873	0.9660
49 50	150	1	0	0.9315	0.0199	0.8797	0.9615
50 51	149	1	0	0.9252	0.0208	0.8721	0.9568
53 54	148	0	1	0.9252	0.0208	0.8721	0.9568
59 60	147	1	0	0.9189	0.0216	0.8645	0.9521
62 63	146	1	1	0.9126	0.0223	0.8569	0.9473
67 68	144	1	0	0.9063	0.0230	0.8493	0.9424
75 76	143	1	0	0.8999	0.0237	0.8418	0.9375
84 85	142	1	0	0.8936	0.0244	0.8344	0.9325
90 91	141	1	0	0.8873	0.0250	0.8270	0.9274
95 96	140	1	0	0.8809	0.0256	0.8197	0.9223
96 97	139	1	0	0.8746	0.0262	0.8124	0.9172
98 99	138	0	1	0.8746	0.0262	0.8124	0.9172
103 104	137	0	1	0.8746	0.0262	0.8124	0.9172
111 112	136	0	1	0.8746	0.0262	0.8124	0.9172
117 118	135	1	0	0.8681	0.0268	0.8049	0.9119
126 127	134	1	0	0.8616	0.0274	0.7975	0.9066
127 128	133	1	0	0.8551	0.0279	0.7901	0.9013
129 130	132	1	1	0.8486	0.0285	0.7827	0.8959
136 137	130	1	0	0.8421	0.0290	0.7753	0.8905
145 146	129	1	0	0.8356	0.0295	0.7679	0.8850

```
more
```

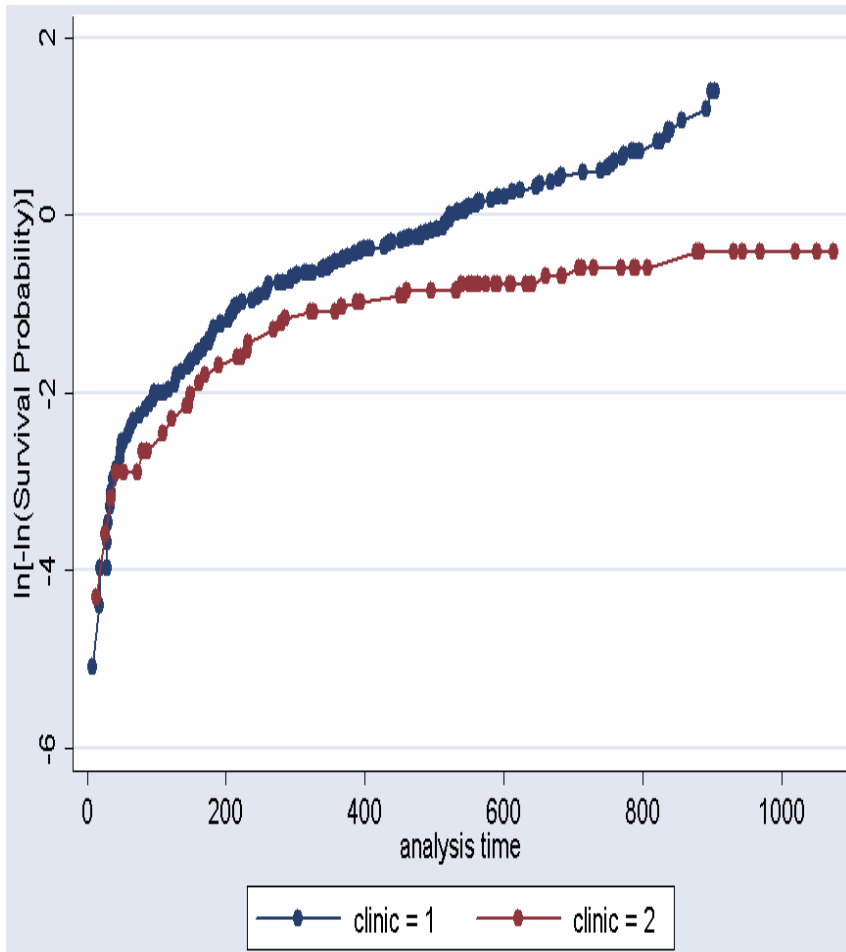
- Ltable survt status, by(clinic)

2. Assessing ph assumption using graphs

- 1) log-log KM survival estimates (stratified by CLINIC) plotted against time (or against the log of time)
 - 2) Log-log Cox adjusted survival estimates (stratified by CLINIC) plotted against time
- * The first two are based on whether log-log survival curves are parallel for different levels of CLINIC (stphplot).
 - KM survival estimates and Cox adjusted survival estimates plotted on the same graph
 - * Are predicted values from PH model from Cox close to observed values using KM? (stcoxkm)

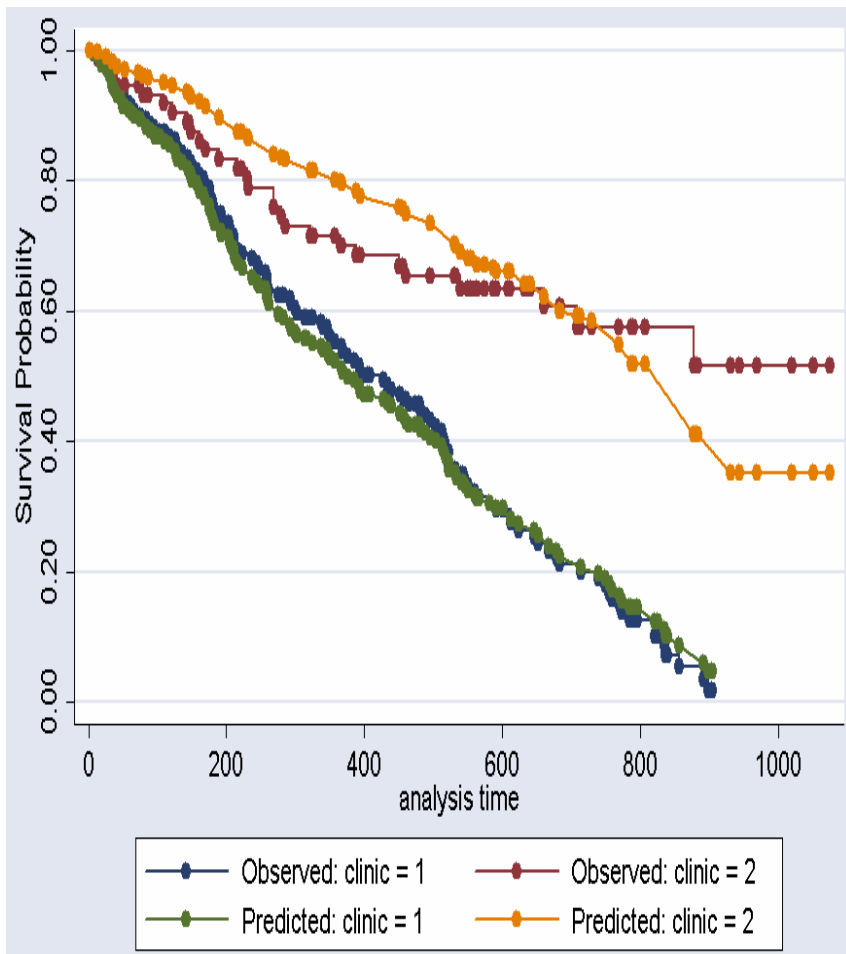


- Staphplot, by(clinic): left picture produced
- Staphplot, by(clinic) nonegative: $\log(-\log)$ plots (upward) rather than the default $-\log(-\log)$ plots (downward)
- Stata (SAS) produced $\log(\text{survival time})$ rather than survival time on the x axis.



- Stphplot, by(clinic) nonegative nolntime
- If the log-log survival curves look like straight lines with $\log(\text{survival time})$ on the x, this means the time-to-event variable follows a Weibull distribution.
- If the slope equals one, survival time variable follows an exponential distribution.
- For both, parametric models can be used.

- `Stpplot, strata(clinic) adjust(prison dose) nonegative nolntime`
- Log-log curves are adjusted for PRISON and DOSE using a stratified Cox model on the variable CLINIC.



- `Stcoxkm, by(clinic)`
- Two curves are very close together for `CLINIC=1` and less for `CLINIC=2`, which suggests there is some violation with PH assumption. The predicted values are Cox adjusted for `CLINIC`, and therefore assume PH assumption.

3. Running Cox PH model

- * The hazard is proportional across different patterns of covariates: the same baseline hazard for all possible patterns of covariates.

```

stcox prison clinic dose, nohr

      failure _d: status == 1
analysis time _t: survt
           id: id

Iteration 0:  log likelihood = -705.6619
Iteration 1:  log likelihood = -674.54907
Iteration 2:  log likelihood = -673.407
Iteration 3:  log likelihood = -673.40242
Iteration 4:  log likelihood = -673.40242
Refining estimates:
Iteration 0:  log likelihood = -673.40242

Cox regression -- Breslow method for ties

No. of subjects =      238           Number of obs =      238
No. of failures =      150
Time at risk   =      95812
Log likelihood = -673.40242           LR chi2(3) =      64.52
                                           Prob > chi2 =      0.0000

+-----+-----+-----+-----+-----+-----+
      _t |      Coef.  Std. Err.      z    P>|z|    [95% Conf. Interval]
+-----+-----+-----+-----+-----+-----+
      prison |      .3265108   .1672211     1.95  0.051   - .0012366   .6542581
      clinic |     -1.00887   .2148709    -4.70  0.000   -1.430009   -.5877304
      dose   |     -.0353962   .0063795    -5.55  0.000   -.0478997   -.0228926
+-----+-----+-----+-----+-----+

```

- Stcox prison clinic dose, nohr
- Nohr option: regression coefficient rather than the default (hazard ratios)
- Do not need to submit time-to-event variable or status variable if you did stset at the beginning.
- Otherwise, cox survt prison clinic dose, dead(status)

```

. stcox prison clinic dose, nohr exactm

      failure _d: status == 1
analysis time _t: survt
           id: id

Iteration 0:  log likelihood = -698.60784
Iteration 1:  log likelihood = -667.47343
Iteration 2:  log likelihood = -666.33197
Iteration 3:  log likelihood = -666.3274
Iteration 4:  log likelihood = -666.3274
Refining estimates:
Iteration 0:  log likelihood = -666.3274

Cox regression -- exact marginal likelihood

No. of subjects =      238      Number of obs   =      238
No. of failures =      150
Time at risk    =     95812
Log likelihood  =    -666.3274      LR chi2(3)    =     64.56
                                           Prob > chi2   =     0.0000


```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Intervals]	
prison	.326581	.1672306	1.95	0.051	-.0011849	.6543469
clinic	-1.009906	.2148906	-4.70	0.000	-1.431084	-.5887285
dose	-.0353694	.0063789	-5.54	0.000	-.0478718	-.0228669

- The default method of handling ties is Breslow.
- Stcox prison clinic dose, nohr exactm: exact marginal likelihood
- Stcox prison clinic dose, nohr exactp: exact partial likelihood
- If there are a lot of events that occur at the same time, exact methods are preferred.

```

. generate clin_pr=clinic*prison
. stcox prison clinic dose clin_pr, nohr

      failure_d:  status == 1
analysis time _t:  survt
           id:  id

Iteration 0:  log likelihood = -705.6619
Iteration 1:  log likelihood = -673.06525
Iteration 2:  log likelihood = -672.07667
Iteration 3:  log likelihood = -672.07448
Refining estimates:
Iteration 0:  log likelihood = -672.07448

Cox regression -- Breslow method for ties

No. of subjects =      238           Number of obs =      238
No. of failures =      150
Time at risk   =     95812
Log likelihood = -672.07448           LR chi2(4) =      67.17
                                           Prob > chi2 =     0.0000

```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
prison	1.163988	.5403366	2.15	0.031	.1049475 2.223028
clinic	-.6541364	.2890204	-2.26	0.024	-1.220606 -.0876668
dose	-.0370059	.006513	-5.68	0.000	-.049771 -.0242407
clin_pr	-.6993311	.4293122	-1.63	0.103	-1.540767 .1421053

- For interaction terms,
- Generate
clin_pr=clinic*prison
- And then, stcox
prison clinic dose
clin_pr, nohr

```
. lrtest
You ran lrtest using the old syntax. Click here to learn about the new syntax.

likelihood-ratio test                                LR chi2(1) =      2.66
(Assumption: . nested in LRTEST_0)                 Prob > chi2 =     0.1032
```

- Lrtest (likelihood ratio test)
- Lrtest, saving(0)
- And then, for reduced model: stcox prison clinic dose
- Lrtest again

4. Running a stratified cox model

```

. stcox prison dose, strata(clinic)

      failure_d: status == 1
      analysis time _t: survt
              id: id

Iteration 0:  log likelihood = -614.68365
Iteration 1:  log likelihood = -597.73516
Iteration 2:  log likelihood =  -597.714
Refining estimates:
Iteration 0:  log likelihood =  -597.714

Stratified Cox regr. -- Breslow method for ties

No. of subjects =      238          Number of obs =      238
No. of failures =      150
Time at risk   =     95812
Log likelihood =  -597.714          LR chi2(2)   =     33.94
                                      Prob > chi2   =     0.0000

+-----+-----+-----+-----+-----+-----+
      _t | Haz. Ratio | Std. Err. | z | P>|z| | [95% Conf. Interval]
+-----+-----+-----+-----+-----+-----+
      prison | 1.475192 | .2491827 | 2.30 | 0.021 | 1.059418 2.054138
      dose   | .9654655 | .0062418 | -5.44 | 0.000 | .953309 .977777
+-----+-----+-----+-----+-----+-----+
                                      Stratified by clinic

```

- If the ph assumption is not met for the variable CLINIC, but is met for the variables PRISON and DOSE, then a stratified Cox analysis can be performed.
- Stcox **prison dose**, strata (clinic)

- How to estimate the hazard ratio for PRISON=1 vs PRISON=0 for CLINIC=2.
- Lincom command can be used to exponentiate linear combinations of parameters: lincom prison+2*clin_pr, hr

$$HR = \frac{h_0(t) \exp[b_1(1) + b_2(DOSE) + b_3(2)(1)]}{h_0(t) \exp[b_1(0) + b_2(DOSE) + b_3(2)(0)]} = \exp(b_1 + 2b_3)$$

```

.stcox prison dose if clinic==2

      failure _d: status == 1
      analysis time _t: survt
              id: id

Iteration 0:  log likelihood = -109.22183
Iteration 1:  log likelihood = -104.38685
Iteration 2:  log likelihood = -104.37135
Iteration 3:  log likelihood = -104.37135
Refining estimates:
Iteration 0:  log likelihood = -104.37135

Cox regression -- Breslow method for ties

No. of subjects =      75      Number of obs =      75
No. of failures =      28
Time at risk   =    36254
Log likelihood = -104.37135      LR chi2(2)   =      9.70
                                      Prob > chi2   =     0.0078

```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
prison	.9210324	.3539571	-0.21	0.831	.4336648 1.956121
dose	.9637452	.0118962	-2.99	0.003	.9407088 .9873457

- Models can also be run on a subset portion of the data using if statement: `stcox prison dose if clinic==2`
- Hazard ratio estimates for PRISON=1 vs PRISON=0 (for CLINIC=2) are exactly the same when using the stratified Cox approach with product terms and the subset data approach.

5. Assessing the PH assumption using a statistical test

- The command `stphtest` gives PH global test for all the covariates simultaneously (Schoenfeld residuals) and can also be used to obtain a test for each covariate separately (Schoenfeld scaled residuals) with the `detail` option.
- If PH assumption is met, the residuals should not be correlated with survival time (or ranked time)
- Before this test, `stcox` with `schoenfeld()` option.

```

. stcox prison dose clinic, schoenfeld(schoen*) scaledsch(scaled*)

      failure _d: status == 1
analysis time _t: survt
              id: id

Iteration 0:  log likelihood = -705.6619
Iteration 1:  log likelihood = -674.54907
Iteration 2:  log likelihood = -673.407
Iteration 3:  log likelihood = -673.40242
Iteration 4:  log likelihood = -673.40242
Refining estimates:
Iteration 0:  log likelihood = -673.40242

Cox regression -- Breslow method for ties

No. of subjects =      238          Number of obs =      238
No. of failures =      150
Time at risk   =      95812

Log likelihood = -673.40242          LR chi2(3) =      64.52
                                      Prob > chi2 =      0.0000

+-----+-----+-----+-----+-----+-----+
      _t | Haz. Ratio | Std. Err. | z | P>|z| | [95% Conf. Interval]
+-----+-----+-----+-----+-----+-----+
      prison | 1.386123 | .231789 | 1.95 | 0.051 | .9987642 | 1.923715
      dose | .965223 | .0061576 | -5.55 | 0.000 | .9532294 | .9773675
      clinic | .3646309 | .0783486 | -4.70 | 0.000 | .2393068 | .5555868
+-----+-----+-----+-----+-----+-----+

```

- Stcox prison dose clinic, schoenfeld(schoen*) scaledsch(scaled*)

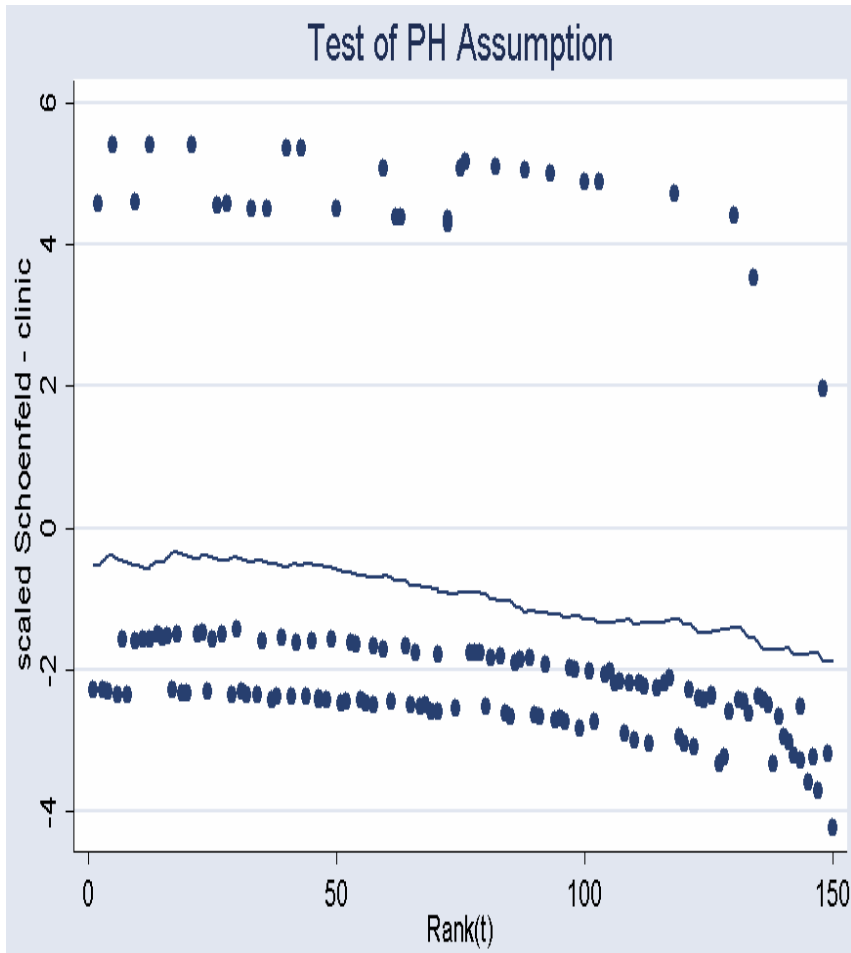
```
. stphtest, rank detail
```

```
Test of proportional hazards assumption
```

```
Time: Rank(t)
```

	rho	chi2	df	Prob>chi2
prison	-0.04645	0.32	1	0.5689
dose	0.08975	1.08	1	0.2996
clinic	-0.24927	10.44	1	0.0012
global test		12.36	3	0.0062

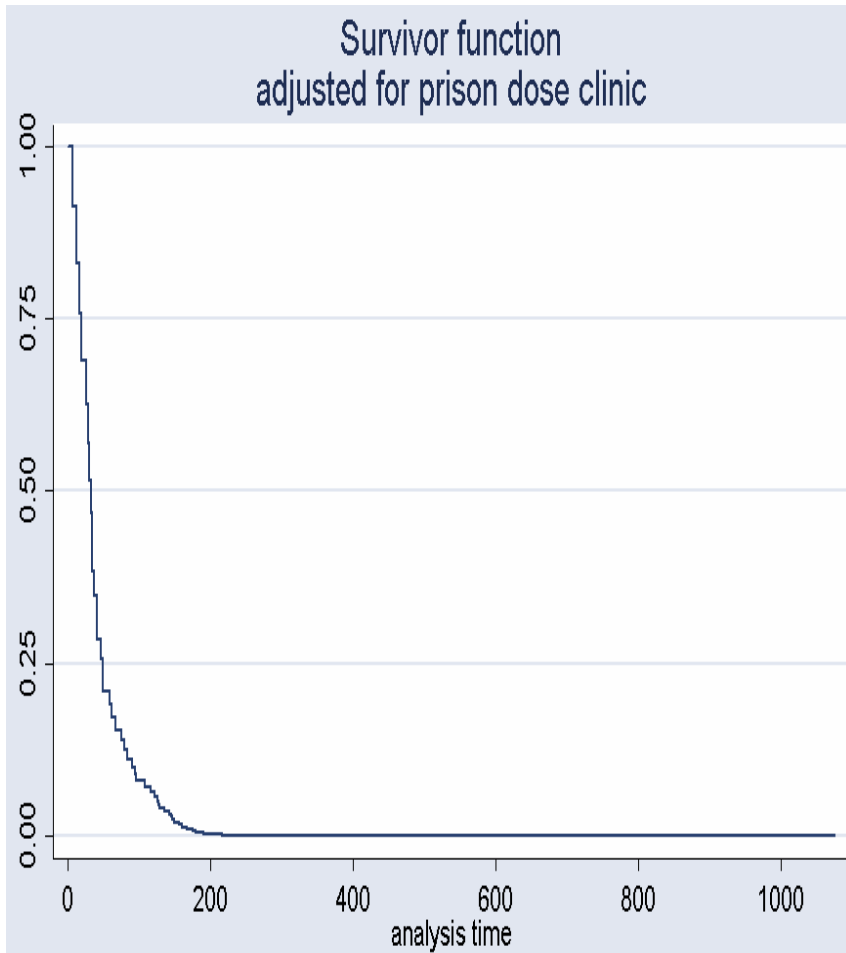
- Stphtest, rank detail
- PH assumption is not met for CLINIC



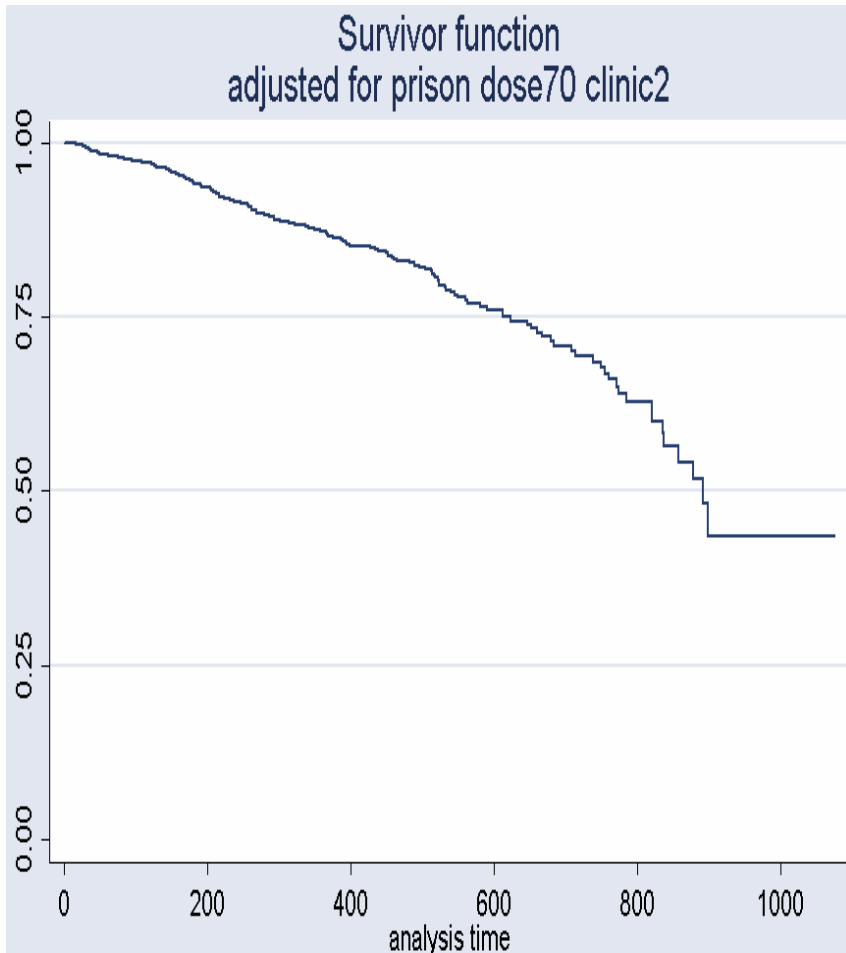
- Stptest, rank plot(clinic): this produces a plot of the scaled Schoenfeld residuals for CLINIC against survival time ranking.

6. Obtaining Cox adjusted survival curves

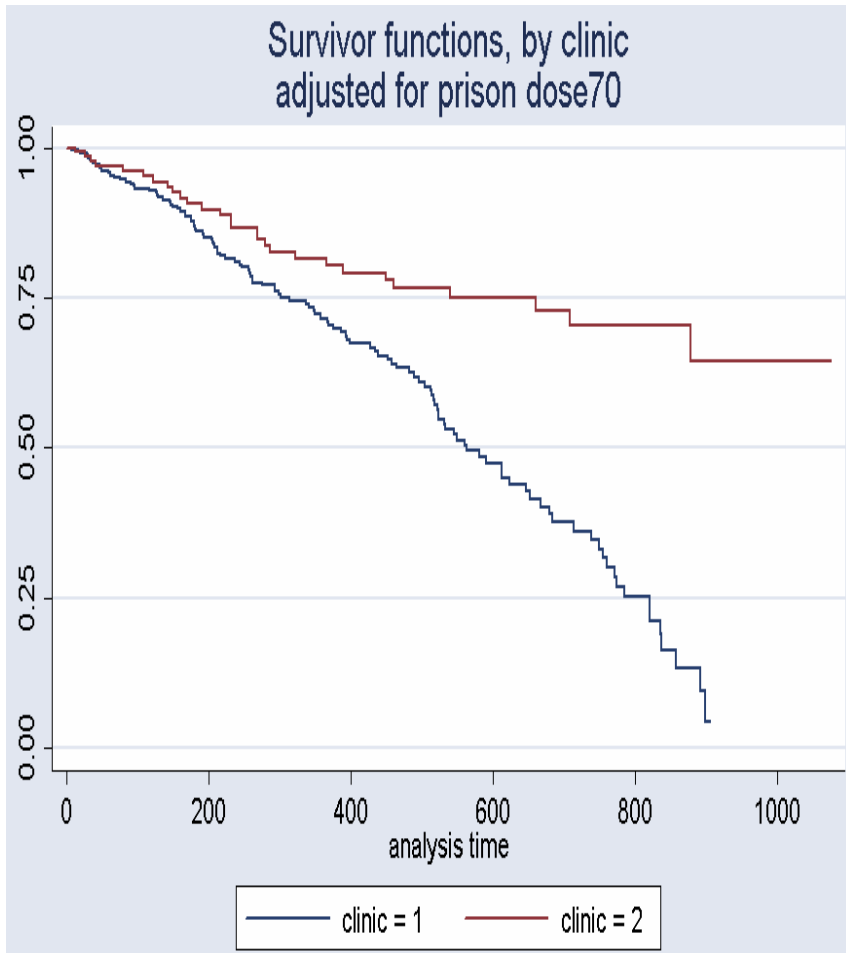
- * Adjusted survival curves depending on the pattern of covariates can be obtained with `sts graph` command.



- Sts graph,
adjustfor(prison dose
clinic): adjusted
survival plot with
PRISON=0,
CLINIC=0, DOSE=0



- If you are interested in graphing that curve for PRISON=0, CLINIC=2, DOSE=70,
- First, generate clinic2=clinic-2; generate dose70=dose-70,
- Sts graph, adjustfor(prison dose70 clinic2)



- Sts graph, `strata(clinic) adjustfor (prison dose70)`
- This curve suggests a strong effect from **CLINIC** on survival.

- Compare adjusted curves of PRISON=1 to PRISON=0 stratified by CLINIC.
- Generate `prison1=prison-1; sts generate scox0=s, strata(clinic) adjustfor(prison dose70); sts generate scox1=s, strata(clinic) adjustfor(prison1 dose70)`
- Graph `scox0 scox1, twoway symbol ([clinic] [clinic]) xlabel(365,730, 1095)`

7. Running an extended Cox model

- If ph assumption is not met, a possible strategy is to run a stratified Cox model
- Another strategy is an extended model with tvc.
- Tvc option: a time varying covariate will be multiplied by a function of time
- Texp option with the variable `_t` representing time for the specification of that function of time.

```

. stcox prison clinic dose, tvc(dose) texp(ln(_t)) nohr

      failure _d:  status == 1
      analysis time _t:  survt

Iteration 0:  log likelihood = -705.6619
Iteration 1:  log likelihood = -673.81352
Iteration 2:  log likelihood = -672.52172
Iteration 3:  log likelihood = -672.51694
Iteration 4:  log likelihood = -672.51694
Refining estimates:
Iteration 0:  log likelihood = -672.51694

Cox regression -- Breslow method for ties

No. of subjects =      238          Number of obs =      238
No. of failures =      150
Time at risk   =      95812

Log likelihood = -672.51694          LR chi2(4) =      66.29
                                      Prob > chi2 =      0.0000

+-----+-----+-----+-----+-----+-----+
      _t |      Coef.  Std. Err.      z    P>|z|     [95% Conf. Intervall]
+-----+-----+-----+-----+-----+-----+
      rh |
      prison |      .3404817   .1674672    2.03   0.042     .012252   .6687113
      clinic |     -1.018682   .215385   -4.73   0.000    -1.440829  -.5965352
      dose   |     -.0824307   .0359866   -2.29   0.022    -.1529631  -.0118982
+-----+-----+-----+-----+-----+-----+
      t  |
      dose   |      .0085751   .0064554    1.33   0.184    -.0040772   .0212274
+-----+-----+-----+-----+-----+-----+

Note: Second equation contains variables that continuously vary with respect to
interacted with current values of ln(_t).

```

- Stcox prison clinic dose, tvc(dose) texp(ln(_t)) nohr, here tvc is DOSE*ln(time)

```

. stcox prison dose clinic, tvc(clinic) texp(_t)=365) nohr

      failure_d: status == 1
      analysis time _t: survt

Iteration 0:  log likelihood = -705.6619
Iteration 1:  log likelihood = -670.1804
Iteration 2:  log likelihood = -668.59639
Iteration 3:  log likelihood = -668.57444
Iteration 4:  log likelihood = -668.57443
Refining estimates:
Iteration 0:  log likelihood = -668.57443

Cox regression -- Breslow method for ties

No. of subjects =          238      Number of obs   =          238
No. of failures =          150
Time at risk   =          95812
Log likelihood = -668.57443      LR chi2(4)     =          74.17
                                      Prob > chi2    =          0.0000

```

	_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
h	prison	.377704	.1684024	2.24	0.025	.0476414 .7077666
	dose	-.0355116	.0064354	-5.52	0.000	-.0481247 -.0228985
	clinic	-.4595628	.2552911	-1.80	0.072	-.959924 .0407985
t	clinic	-1.368665	.4613948	-2.97	0.003	-2.272982 -.464348

Note: Second equation contains variables that continuously vary with respect to interacted with current values of _t)=365.

- Heavyside function can also be used.
- Stcox prison dose clinic, tvc(clinic) texp(_t)>=365) nohr
- Texp option can be used only once in the stcox command. Not good for two heavyside functions. But,

	id	clinic	status	survt	_st	_d	_t	_t0	v1
1	1	1	.	365	1	0	365	0	0
2	1	1	1	428	1	1	428	365	365
3	2	1	1	275	1	1	275	0	0
4	3	1	1	262	1	1	262	0	0
5	4	1	1	183	1	1	183	0	0
6	5	1	1	259	1	1	259	0	0
7	6	1	.	365	1	0	365	0	0
8	6	1	1	714	1	1	714	365	365
9	7	1	.	365	1	0	365	0	0
10	7	1	1	438	1	1	438	365	365
11	8	1	.	365	1	0	365	0	0
12	8	1	0	796	1	0	796	365	365
13	9	1	.	365	1	0	365	0	0
14	9	1	1	892	1	1	892	365	365
15	10	1	.	365	1	0	365	0	0
16	10	1	1	393	1	1	393	365	365
17	11	1	0	161	1	0	161	0	0
18	12	1	.	365	1	0	365	0	0
19	12	1	1	836	1	1	836	365	365

- Using stsplot, any subject followed more than 365 days is represented by two observations rather than one.

id	_t0	_t	clinic	v1	hv1	hv2
1	0	365	1	0	1	0
1	365	428	1	365	0	1
2	0	275	1	0	1	0
3	0	262	1	0	1	0
4	0	183	1	0	1	0
5	0	259	1	0	1	0
6	0	365	1	0	1	0
6	365	714	1	365	0	1
7	0	365	1	0	1	0
7	365	438	1	365	0	1
8	0	365	1	0	1	0
8	365	796	1	365	0	1
9	0	365	1	0	1	0
9	365	892	1	365	0	1
10	0	365	1	0	1	0
10	365	393	1	365	0	1
11	0	161	1	0	1	0
12	0	365	1	0	1	0
12	365	836	1	365	0	1
13	0	365	1	0	1	0

- With the data in this form, two heavyfunctions can actually be defined in the data using:
- Generate
 $hv2 = clinic * (v1 / 365);$
generate $hv1 = clinic * (1 - (v1 / 365));$ list id _t0 _t clinic v1 hv1 hv2, noobs

```

x regression -- Breslow method for ties

. of subjects =      238      Number of obs =      360
. of failures =      150
Time at risk =     95812

Log likelihood = -668.57443      LR chi2(4) =      74.17
                                Prob > chi2 =      0.0000

```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
prison	.377704	.1684024	2.24	0.025	.0476414	.7077666
dose	-.0355116	.0064354	-5.52	0.000	-.0481247	-.0228985
hv1	-.4595628	.2552911	-1.80	0.072	-.959924	.0407985
hv2	-1.828228	.385946	-4.74	0.000	-2.584668	-1.071788

- With the two heavyfunctions above, we can use a time-dependent model as follows:
- Stcox prison clinic dose hv1 hv2, nohr
- It is possible to split the data at every single failure time, but a large amount of memory.

- Skip 8. Running Parametric Models and 9. Running Frailty Models

10. Modeling Recurrent Events

* bladder.dta

- Recurrent events are represented in the data with multiple observations for subjects having multiple events.
- Cox model with recurrent events using the counting process approach can be run with the `stcox` command

bladder.dta

. list

	id	event	interval	start	stop	tx	num	size
1.	1	0	1	0	0	0	1	1
2.	2	0	1	0	1	0	1	3
3.	3	0	1	0	4	0	2	1
4.	4	0	1	0	7	0	1	1
5.	5	0	1	0	10	0	5	1
6.	6	1	1	0	6	0	4	1
7.	6	0	2	6	10	0	4	1
8.	7	0	1	0	14	0	1	1
9.	8	0	1	0	18	0	1	1
10.	9	1	1	0	5	0	1	3
11.	9	0	2	5	18	0	1	3
12.	10	1	1	0	12	0	1	1
13.	10	1	2	12	16	0	1	1
14.	10	0	3	16	18	0	1	1
15.	11	0	1	0	23	0	3	3
16.	12	1	1	0	10	0	1	3
17.	12	1	2	10	15	0	1	3
18.	12	0	3	15	23	0	1	3
19.	13	1	1	0	3	0	1	1
20.	13	1	2	3	16	0	1	1

```

. stset stop, failure (event==1) id(id) time0(start) exit(time .)

      id:      id
failure event: event == 1
obs. time interval: (start, stop]
exit on or before: time .

```

191	total obs.	
1	entry on or after exit (start>stop)	PROBABLE ERROR

190	obs. remaining, representing	
85	subjects	
112	failures in multiple failure-per-subject data	
2711	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	64

- stset stop, failure (event==1) id(id) time0(start) exit(time .)
- Id option for clustered variables
- Time0() defines the variable that begins the time interval
- Exit(time .) defined no imposed limit on the length of follow-up time for a given subject (e.g. subjects are not out of the risk set after their first event)

```
. list id _t0 _t _d tx in 12/20
```

	id	_t0	_t	_d	tx
12.	10	0	12	1	0
13.	10	12	16	1	0
14.	10	16	18	0	0
15.	11	0	23	0	0
16.	12	0	10	1	0
17.	12	10	15	1	0
18.	12	15	23	0	0
19.	13	0	3	1	0
20.	13	3	16	1	0

- List id _t0 _t _d tx in 12/20

```

Cox regression -- Breslow method for ties
No. of subjects      =           85      Number of obs      =           190
No. of failures      =           112
Time at risk         =           2711
Log pseudo-likelihood = -460.07958      Wald chi2(3)       =           11.25
                                                Prob > chi2        =           0.0105

                                <standard errors adjusted for clustering on id>

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
tx	-.4070966	.2432658	-1.67	0.094	-.8838889	.0696956
num	.1606478	.0572305	2.81	0.005	.0484781	.2728174
size	-.0400877	.0726459	-0.55	0.581	-.182471	.1022957

- Stcox tx num size, nohr robust
- Robust option requests robust standard errors for the coefficient estimates.

```

Stratified Cox regr. -- Breslow method for ties
No. of subjects      =           85          Number of obs      =          190
No. of failures      =           112
Time at risk        =          2711
Log pseudo-likelihood = -319.85912          Wald chi2(3)       =           7.11
                                                Prob > chi2        =           0.0685
                                                <standard errors adjusted for clustering on id>

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
tx	-.3342955	.1982339	-1.69	0.092	-.7228268 .0542359
num	.1156526	.0502089	2.30	0.021	.017245 .2140603
size	-.0080508	.0604807	-0.13	0.894	-.1265908 .1104892

Stratified by interval

- A stratified model can be run using the data with the variable INTERVAL. This stratified variable indicates whether subjects were at risk for their 1st, 2nd, 3rd, or 4th event. This approach called “Conditional 1” to distinguish the order in which recurrent events occur.
- **stcox tx num size, nohr robust strata(interval)**

- Interaction terms between TX(treatment) and the stratified variable could be created to examine whether the effect of treatment differed for the 1st, 2nd, 3rd, or 4th event.
- But, how?
- Interval dummy needed? (INTD)
- And, generate TX*INTD
- And, `stcox tx num size tx*ind, nohr robust?`
- Or, generate TX*INT
- And, `stcox tx num size tx*int, nohr strata(interval)?`

```
. list id _t0 _t _d tx in 12/20
```

	id	_t0	_t	_d	tx
12.	10	0	12	1	0
13.	10	0	4	1	0
14.	10	0	2	0	0
15.	11	0	23	0	0
16.	12	0	10	1	0
17.	12	0	5	1	0
18.	12	0	8	0	0
19.	13	0	3	1	0
20.	13	0	13	1	0

- In another approach (conditional 2), the starting time at risk gets reset to zero for each subsequent event. (The difference between 1 and 2 is in the way the time intervals for recurrent events are defined)
- Generate stop2=_t-t0
- Stset stop2, failure (event==1) exit (time .)

```

Stratified Cox regr. -- Breslow method for ties
No. of subjects      =          190          Number of obs      =          190
No. of failures      =          112
Time at risk        =          2711
Log pseudo-likelihood = -363.16022
Wald chi2(3)        =          11.99
Prob > chi2         =          0.0074

                (standard errors adjusted for clustering on id)

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
tx	-.2695213	.2093108	-1.29	0.198	-.6797628	.1407203
num	.1535334	.0491803	3.12	0.002	.0571418	.2499249
size	.0068402	.0625862	0.11	0.913	-.1158265	.129507

Stratified by interval

- Notice that id option was not used in this conditional approach 2. This means that Stata does not know that multiple observations correspond to the same subject. However, the cluster() option can be employed to request that the analysis be clustered by id.
- Stcox tx num size, nohr robust strata(interval) cluster(id)

```

. stset stop, failure (event==1) id(id) time0(start) exit(time .)

      id:      id
failure event: event == 1
obs. time interval: (start, stop]
exit on or before: time .

```

191	total obs.	
1	entry on or after exit (start>stop)	PROBABLE ERROR

190	obs. remaining, representing	
85	subjects	
112	failures in multiple failure-per-subject data	
2711	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	64

- How a shared frailty model can be applied to recurrent event data. Frailty is included in recurrent event analyses to account for variability due to unobserved subject specific factors that may lead to within-subject correlation.
- Make the data the original format

```

Weibull regression --
    log-relative hazard form          Number of obs      =      190
    Gamma shared frailty              Number of groups    =       85
Group variable: id

No. of subjects =          85          Obs per group: min =          1
No. of failures =          112         avg   =      2.235294
Time at risk    =          2711        max   =          5

Log likelihood =      -184.73658      LR chi2(3)          =          8.04
                                          Prob > chi2         =          0.0453

```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Intervall
tx	-.4583219	.2677275	-1.71	0.087	-.9830582 .0664143
num	.1847305	.0724134	2.55	0.011	.0428028 .3266581
size	-.0314314	.0911134	-0.34	0.730	-.2100104 .1471476
_cons	-2.952397	.4174276	-7.07	0.000	-3.77054 -2.134254
/ln_p	-.1193215	.0898301	-1.33	0.184	-.2953852 .0567421
/ln_theta	-.7252604	.5163027	-1.40	0.160	-1.737195 .2866742
p	.8875224	.0797262			.7442449 1.058383
1/p	1.126732	.1012144			.9448377 1.343644
theta	.4841985	.249993			.1760134 1.33199

```

Likelihood-ratio test of theta=0: chibar2(01) =      7.34 Prob>=chibar2 = 0.003

```

- A parametric Weibull model is run with a gamma distributed shared frailty component using the streg command.
- Streg tx num size, dist(weibull) frailty(gamma) shared(id) nohr
- Shared option defines the cluster variable. In this model, observations from the same subject share the same frailty.