

Competing-risks regression

Roberto G. Gutierrez

Director of Statistics
StataCorp LP

Stata Conference Boston 2010

1. Overview
2. The problem of competing risks
3. Cause-specific hazards
4. Cumulative incidence functions
5. The Cox regression approach
6. The Fine and Gray approach (`stcrreg`)
7. Concluding remarks

- Competing risks are events that occur instead of the failure event of interest, and we cannot treat these as censored
- When you have competing events, you want to focus on *cause-specific* hazards rather than standard hazards
- When you have competing events, you want to focus on the *cumulative incidence function* (CIF) rather than the survival function
- Cox regression is fine for cause-specific hazards, but for CIFs you need to go through a lot of work
- Competing-risks regression by the method of Fine and Gray (1999) is a useful alternative
- Implemented in the `stcrreg` command, new to Stata 11

- A competing-risk event is an event that impedes what a researcher actually “wants” to see
- For example, if a researcher is interested in recurrence of breast cancer, cancer occurring in another location would be a competing event
- In general you cannot treat competing events as censored because
 1. The competing events might be dependent, and you usually can't test this
 2. You are unwilling to apply your results to a counter-factual world where the competing event doesn't exist

- What confuses the matter is that, often, it is okay to think of competing events as censorings. But this is merely a computational device
- That is, the software will often let you treat competing events as censored, but your interpretation of the results should never do so
- Even more confusing is that some software, such as that for Kaplan-Meier curves, is not appropriate if you treat competing events as censored

- Consider data on 423 women treated for breast cancer
- 175 women were given an experimental new drug, and the other 248 women the standard therapy
- Time in months until disease progression occurred either at an old tumor site (local relapse) or at a new site (distant relapse) was recorded; local relapse is the event of interest
- If no event occurred after within 60 months, observations were censored
- Age and race also recorded

```
. use bc_compete, clear
(Breast cancer with competing risks)
```

```
. describe
```

```
Contains data from bc_compete.dta
```

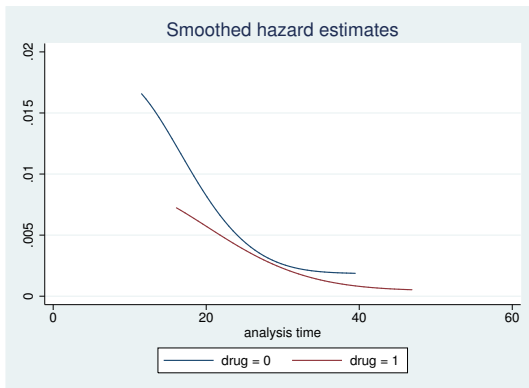
```
  obs:          423                      Breast cancer with competing risks
 vars:           5                      13 Jul 2010 04:27
 size:        11,844 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
age	float	%9.0g		Age at start of the trial
drug	float	%9.0g		1 treatment; 0 control
race	float	%9.0g	race	1 white; 2 black; 3 other
time	float	%9.0g		time in months
status	float	%9.0g		0 censored; 1 local relapse; 2 distant relapse

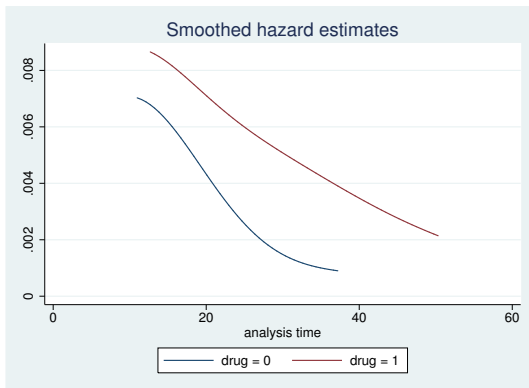
```
Sorted by:
```

- A cause-specific hazard is the instantaneous risk of failure from a specific cause given that failure (from any cause) has yet to happen
- In our example, we have two cause-specific hazards: one for local relapse and one for distant relapse
- You can estimate, graph, and test a cause-specific hazard with “standard” survival software, if you treat the other event as censored
- You need only to modify your interpretation, specifically, you need to consider both hazards jointly and not attempt to disentangle them

```
. stset time, failure(status = 1)  
  (output omitted)  
. sts graph, hazard by(drug) kernel(gauss)
```



```
. stset time, failure(status = 2)  
  (output omitted)  
. sts graph, hazard by(drug) kernel(gauss)
```



Because we believe in proportionality of cause-specific hazards we can use `stcox` to test for differences

```
. quietly stset time, failure(status = 1)
. stcox drug
Cox regression -- Breslow method for ties
No. of subjects =          423                Number of obs   =          423
No. of failures =          135
Time at risk    =          15018
Log likelihood   =   -785.18921                LR chi2(1)         =          3.85
                                                Prob > chi2        =          0.0496
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	.7030953	.1283705	-1.93	0.054	.4915895	1.005601

```
. quietly stset time, failure(status = 2)
. stcox drug
Cox regression -- Breslow method for ties
No. of subjects =          423                Number of obs   =          423
No. of failures =           78
Time at risk    =        15018
Log likelihood  =   -443.31893                LR chi2(1)        =          6.75
                                                Prob > chi2        =         0.0094
```

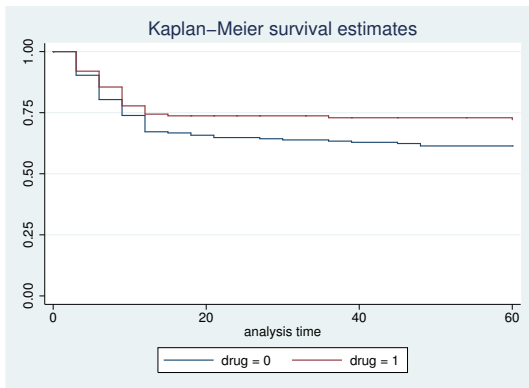
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	1.804488	.4120596	2.58	0.010	1.1534	2.823112

Interpretation: Drug treatment somewhat decreases risk of local relapse while at the same time increasing the risk of distant relapse

- Despite the ominous name, a CIF is just the probability that you observe a specific type of event before a given time
- In our analysis, we have two CIFs; one for local relapse and one for distant relapse
- For example, the CIF for local relapse at 5 months is just the probability of a local relapse before 5 months
- CIFs begin at zero at time zero and increase to an upper limit equal to the eventual probability that the event will take place, but this is not equal to one because of competing events
- Mathematically, the CIF for local relapse is a function of *both* cause-specific hazards

Kaplan-Meier curves are not appropriate for competing risks

- . quietly stset time, failure(status = 1)
- . sts graph, survival by(drug)



- The previous Kaplan-Meier curve attempts to answer the question "What is the probability of no local relapse before (say) 5 months?"
- In a competing-risks setting, a Kaplan-Meier curve is inadequate for three reasons
- First, it fails to acknowledge that local relapse may never occur. In reality, the probability of local relapse after time zero is not equal to 1
- Second, the Kaplan-Meier calculation does not take into account dependence between competing events

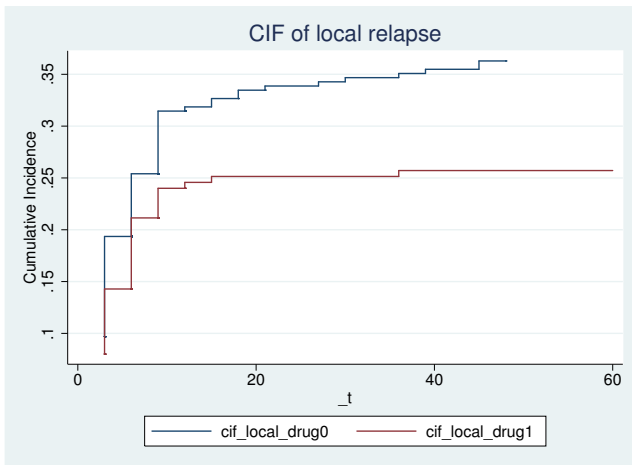
- Third, with competing risks it is better to reverse the temporal ordering of the question
- It makes better sense to ask “What is the probability of local relapse within 5 months?” than to ask “What is the probability that nothing happens for the first five months, but when something does happen I want it to be a local and not a distant relapse?”
- In summary, using Kaplan-Meier demands too much of your data; it requires independent risks and a world where the competing event doesn't occur
- As such, you should use the cumulative incidence function (CIF) instead

You can estimate CIFs nonparametrically using `stcompet`, by Coviello and Boggess (2004) and available from the SSC

```
. quietly stset time, failure(status = 1)
. stcompet cif = ci, compet1(2) by(drug)

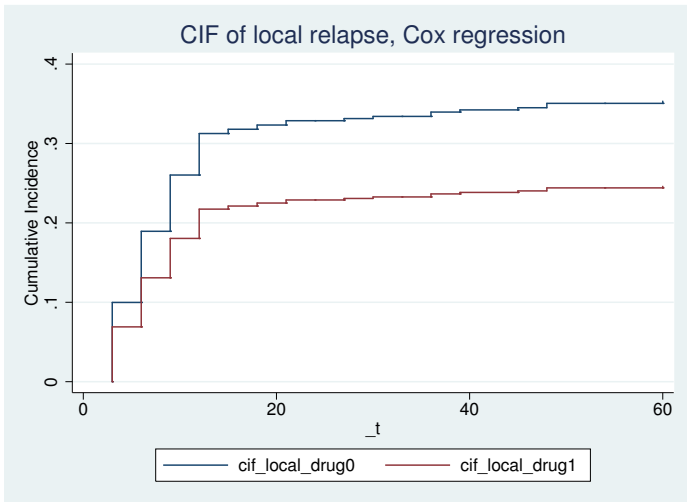
. gen cif_local_drug0 = cif if status == 1 & drug == 0
(333 missing values generated)
. gen cif_local_drug1 = cif if status == 1 & drug == 1
(378 missing values generated)

. twoway line cif_local*_t, connect(step step) sort
> ytitle(Cumulative Incidence) title(CIF of local relapse)
```



- Nonparametric estimation is flexible, but it cannot adjust for external covariates such as age and race
- Previously we applied Cox regression on drug to both cause-specific hazards
- These we could then extend to adjust for age and race, at the cost of the proportionality assumption
- We could then use the resulting cause-specific hazards to derive estimates of the CIFs

- To calculate a CIF from Cox regression, we need to
 1. Predict the baseline hazard contributions from both Cox regressions on drug, that for local relapse and that for distant relapse
 2. Transform the baseline hazard contributions (which assume $\text{drug} == 0$) to those for $\text{drug} == 1$ where appropriate
 3. Use the hazard contributions to calculate a product limit estimator of event-free survival for both levels of drug
 4. Calculate the estimated CIF manually for both $\text{drug} == 0$ and $\text{drug} == 1$; see page 209 of [ST] for details
 5. Plot the results
- The point: Assessing covariate effects on the CIF using Cox regression is a lot of work



- An easier way to do CIF covariate analysis is with competing risks regression, according to the model of Fine and Gray (1999)
- They posit a model for the hazard of the subdistribution for the failure event of interest, known as the *subhazard*
- Unlike cause-specific hazards, there is a one-to-one correspondence between subhazards and CIFs for respective event types; that is the CIF for local relapse is a function of *only* the subhazard for local relapse
- Covariates affect the subhazard proportionally, similar to Cox regression
- You do this in Stata 11 using `stcrreg`. `stcurve` after `stcrreg` will plot comparative CIFs for you

```
. quietly stset time, failure(status = 1)
```

```
. stcrreg drug, compete(status = 2)
```

```
Competing-risks regression
```

```
No. of obs      =      423
```

```
No. of subjects =      423
```

```
Failure event   : status == 1
```

```
No. failed      =      135
```

```
Competing event: status == 2
```

```
No. competing   =       78
```

```
No. censored    =      210
```

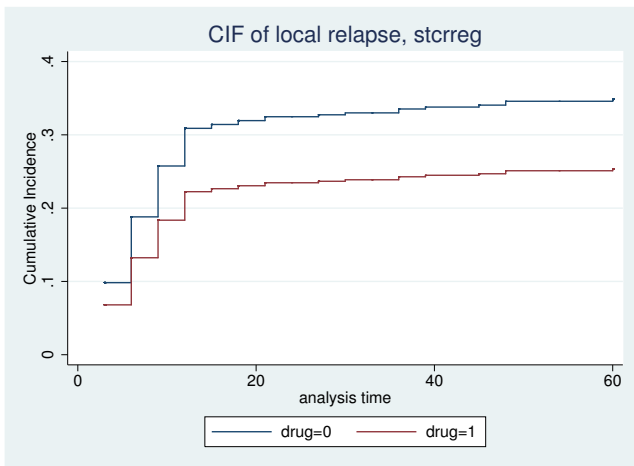
```
Wald chi2(1)    =       4.68
```

```
Log pseudolikelihood = -794.95545
```

```
Prob > chi2     =     0.0305
```

_t	Robust		z	P> z	[95% Conf. Interval]	
	SHR	Std. Err.				
drug	.6810812	.1209317	-2.16	0.031	.4809068	.9645769

```
. stcurve, cif at1(drug=0) at2(drug=1) title("CIF of local relapse, stcrreg")
```



- Competing risks are events that prevent an event of interest from occurring
- If you have competing risks, you want to look at cause-specific hazards instead of standard hazards
- If you have competing risks, you want to look at CIFs instead of survival functions
- CIF analysis with Cox regression is possible, but difficult
- `stcrreg` followed by `stcurve` is the easier way to go
- Keep in mind, however, that easier does not mean correct. There are model assumptions to be made for either of the two approaches

- Coviello, V. and M. Boggess. 2004. Cumulative incidence estimation in the presence of competing risks. *The Stata Journal*. **4**: 103–112.
- Fine, J. and R. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. **94**: 496–509.