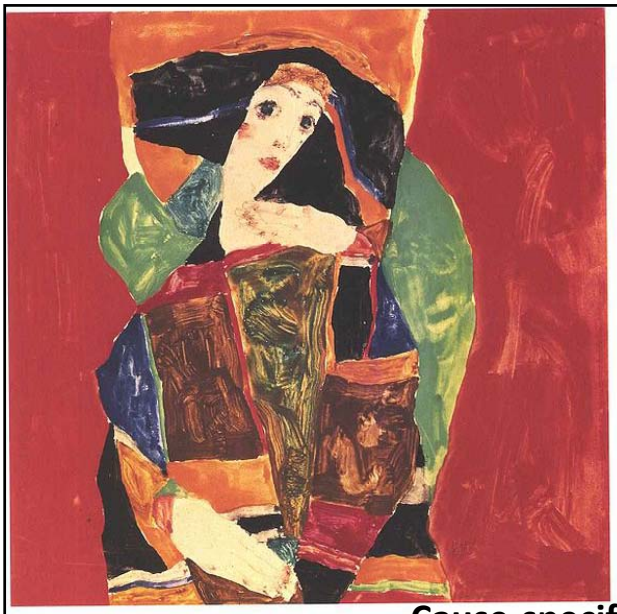
Outline

- Cause-specific survival models in the setting of competing risks
- Data augmentation method
  - Preparation of data
  - Estimate of survival regression models on expanded data
- Covariate-adjusted cumulative incidence: `stcompadj`
- Comparing covariate-adjusted cumulative incidence obtained by two approaches: Cox model (`-stcompadj-`) vs. Fine and Gray method (`-stcrreg- post-estimation`)

2



Cause-specific survival models in the setting of competing risks

3

Common survival study Vs. **Competing risks survival study**

Common survival study measures the time from a starting point to an endpoint defined by the occurrence of one type of event.

In the competing risks situation more than one type of event may occur. Often, one type of event is singled out as the event of interest, the other types as competing events. The remarkable feature of this situation is that the occurrence of one type of event either precludes or fundamentally alters the probability of occurrence of the others[(1)].

4

Hazard - Probability of Occurrence

The hazard is the fundamental measure of the occurrence of the event and we can investigate the role of various factors in modifying the hazard by using appropriate regression models

In the competing risks situation we can still compute the cause-specific hazard for the event of interest and for the competing event. The "latent failure times" mathematical approach demonstrates that the cause-specific hazard is an estimable function also in a competing risks situation and, thus, we can model it[(2)].

5

Hazard - Probability of Occurrence

The probability that the event occurs before time t can be derived from the hazard through an equation. So, the hazard completely describes this probability distribution. Higher the hazard, higher the probability that the event occurs before t and vice versa.

In a competing risks situation, the probability that the main event occurs before time t (Cumulative Incidence) depends on both the hazard of the main event and the hazard of the competing event. Thus, there is no obvious relationship between the hazard and the cumulative incidence of the main event, the latter depending on the hazard of the competing event too.

6

Consequence for the Information

The hazard and the probability that the event occurs before time t convey the same piece of information.

In the competing risks situation the cause-specific hazard and the cumulative incidence do not convey the same piece of information.

The former tells us about the biological mechanism underlying the specific outcome.

The latter informs us about the probability and, therefore, the actual number of patients failing from a specific cause, taking into account that this type of event could not be observed because hindered or precluded from another type of event[2].

7

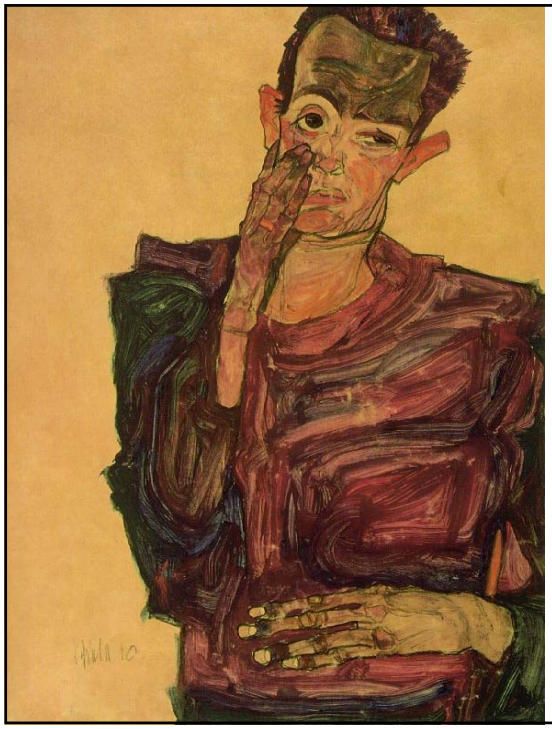
Concluding …

In a competing risks situation, the cause-specific hazard is an estimable function and we can model it by using, for example, a Cox model.

The parameter estimates we obtain

- **are relevant for investigating the factors causing or associated to the occurrence of a specific event**
- **are not appropriate for detecting whether such factors actually change the probability that the event occurs.**

8



Data augmentation method

9

- In a competing risks situation, Cox regression is usually performed by fitting separate models for each cause of failure.
- The hazard corresponding to a specific failure is analyzed considering failures of other causes as censored observations.
- It is possible to make single analysis of two or more cause of failure at the same time rather than fitting separate regression models.
- To perform this analysis a new expanded dataset must be created as follows[3].

10

Here are three subjects from the file malignantmelanoma.dta containing 205 observations:

id	time	thick	sex	cause
51	1516	-.34	M	Malignant Melanoma
52	1525	-1.63	F	Other Causes
53	1542	-2.76	F	Alive

This is a competing risks situation because two causes of failures are present:

1. Death from malignant melanoma (coded as 1)
2. Death from other causes (coded as 2).

Two covariates are present: sex and tumor thickness (centered on its mean)

11

Expanded format

id	thick	sex	cause	stratum	_d	_t
51	-.34	M	Malignant Melanoma	1	1	1516
51	-.34	M	Malignant Melanoma	2	0	1516
52	-1.63	F	Other Causes	1	0	1525
52	-1.63	F	Other Causes	2	1	1525
53	-2.76	F	Alive	1	0	1542
53	-2.76	F	Alive	2	0	1542

- Each subject is represented twice, one for each cause of failure
- A numeric stratum indicator has been created (`stratum`) taking on the value 1 for the first record and value 2 for the second record of the same subject. This allows that a subject has a specific record for each cause of failure
- The failure indicator _d attains the value 1 for each observation of
  - first cause of death in the stratum 1
  - second cause of death in the stratum 2

  and 0 otherwise.

12

In the original data format we can model the effects of gender and thickness of the tumor on the two cause-specific hazards by fitting two separate models:

```
stset time,failure(cause==1)
stcox sex thick
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex	1.775555	.4709861	2.16	0.030	1.055707	2.986242
thick	1.172454	.0383144	4.87	0.000	1.099714	1.250006

```
stset time,failure(cause==2) id(id)
stcox sex thick, nolog
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex	1.720124	.9280661	1.01	0.315	.5974581	4.952361
thick	1.10589	.0847623	1.31	0.189	.9516358	1.285149

13

In the expanded data format we can obtain the same results by fitting a Cox model stratified by the -stratum- indicator variable:

```
stcox stratum#(c.sex c.thick), nolog strata(stratum)
```

```
Stratified Cox regr. -- no ties

No. of subjects =          410                     Number of obs    =       410
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stratum# c.sex						
1	1.775555	.4709861	2.16	0.030	1.055707	2.986242
2	1.720124	.9280661	1.01	0.315	.5974581	4.952361
stratum# c.thick						
1	1.172454	.0383144	4.87	0.000	1.099714	1.250006
2	1.10589	.0847623	1.31	0.189	.9516358	1.285149

Stratified by stratum

14

In the expanded data format we can obtain the same results by fitting a Cox model stratified by the -stratum- indicator variable:

```
stcox stratum#(c.sex c.tick), nolog strata(stratum)
```

The model includes an interaction term between `stratum` and the other covariates

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stratum# c.sex						
1	1.775555	.4709861	2.16	0.030	1.055707	2.986242
2	1.720124	.9280661	1.01	0.315	.5974581	4.952361
stratum# c.thick						
1	1.172454	.0383144	4.87	0.000	1.099714	1.250006
2	1.10589	.0847623	1.31	0.189	.9516358	1.285149

Stratified by stratum

15

Advantages of the expanded data format

- It is possible to fit models where some covariates have exactly the same effect on the main and on the competing event. This cannot be accomplished by fitting separate Cox regressions for each cause of failure[4].
- A major advantage is that we have at hand the baseline hazard contributions of both events (hazard contributions represent the increment of the cumulative hazard at each event time and are the key quantities for estimating the cumulative incidence function).
- The adjustment for covariates is achieved by multiplying the hazard contributions by the exponentiated linear predictor[4].

16

Covariate-adjusted cumulative incidence:

-stcompadj-

17

-stcompadj- is a new Stata macro (v. 10) that computes the adjusted cumulative incidence in the presence of competing risks.

The basic syntax is:

```
stcompadj var [= # var ...] [if] [in] , compet(# [...] )
        [ maineffect(varlist) competeffect(varlist) … ]
```

Fundamentally, -stcompadj- :

- expands the dataset as previously described and fits a Cox or a flexible parametric model whose covariates are the variables specified in `var [[= #] var ...]`
- computes the cumulative incidence function from the baseline hazard contributions and the linear predictor
- saves in two variables the cumulative incidence function for the main and competing event adjusted to the mean or to the specified number of each covariate if the `=#` part is specified.

18

In the following example we estimate the cumulative incidence for a male with a tumor thickness of 2 mm above the mean.

```
stset time,failure(cause==1)
stcompadj sex=1 thick=2, compet(2) maineffect(thick)
```

- compet(#) is not an option. compet(2) means that a failure from a competing event occurs whenever cause takes on the value 2.
- By default the fitted model considers the covariates as having the same effect on the main as well as on the competing event. In the example sex is assumed to have such effect.

19

- The options maineffect(varlist) and competeffect(varlist) allow to fit a model where some of the previously stated variables acts only on the main or on the competing event. In the example thick acts only on the main effect.
- Note that the same variable can be specified both in maineffect(varlist) and in competeffect(varlist). Then, this variable is assumed to have different effects on the main and competing event.
- -stcompadj- creates two variables saving the cumulative incidences functions for the main and competing event. Default names are CI_Main and CI_Compet.

20

The estimates produced by -stcompadj- match the results published by Rosthøj and coll. using the CumInc SAS macro[4]

stcompadj

_t	CI_Main	CI_Compet
10	0.00000	0.006703
30	0.00000	0.013406
99	0.00000	0.020155
185	0.00766	0.020155
204	0.01545	0.020155
210	0.02323	0.020155
232	0.03102	0.026862
232	0.03102	0.026862
279	0.03895	0.026862
295	0.04687	0.026862

CumInc

TIME	P01	P02
10	0.00000	0.006703
30	0.00000	0.013406
99	0.00000	0.020155
185	0.00766	0.020155
204	0.01545	0.020155
210	0.02323	0.020155
232	0.03102	0.026862
279	0.03895	0.026862
295	0.04687	0.026862

21

Further validation of the results produced by -stcompadj- has been done by using the R-macro (si.R and tutfuncs.R) prepared by Putter and coll. for a tutorial in competing risks analysis[5]. Using the si.dta dataset (see slide 37) the results produced by -stcompadj- and R.macros are in very good agreement:

	stcompadj	**R**	**stcompadj**	**R**
time	CI_ccr1_SI	R_ccr1_SI	CI_ccr0_SI	R_ccr0_SI
.112	.0025065	.0025065	.003232	.003232
.137	.0050148	.0050148	.0064639	.0064639
.474	.007525	.007525	.0096959	.0096959
.824	.0100618	.0100618	.0129598	.0129598
.884	.0125986	.0125986	.0162213	.0162213
.....				
12.4	.3784347	.3784347	.3660669	.3660669
12.936	.3872313	.3872313	.3705146	.3705146
13.361	.3872313	.3872313	.3705146	.3705146
13.361	.3872313	.3872313	.3705146	.3705146
13.936	.7015418	.7015418	.5211416	.5211416

22

	stcompadj	R	stcompadj	R
time	CI_ccr1_AIDS	R_ccr1_AIDS	CI_ccr0_AIDS	R_ccr0_AIDS
1.205	0	0	0	0
1.44	.0010797	.0010797	.0036907	.0036907
1.837	.0021758	.0021758	.0074185	.0074185
1.889	.0021758	.0021758	.0074185	.0074185
2.048	.0021758	.0021758	.0074185	.0074185
.....				
11.387	.1927379	.1927379	.4586331	.4586331
11.943	.1962167	.1962167	.4635909	.4635909
12.936	.1998003	.1998003	.4685524	.4685524
13.936	.2074857	.2074858	.4788584	.4788584

23

By specifiying -showmod- we can see the model fitted by -stcompadj- before estimating the cumulative incidence function. For example, the model in the slide 13 is shown by -stcompadj- as follows:

```
stcompadj sex=1 thick=2, maineffect(sex thick) competeffect(sex thick) ///
                         compet(2) showmod
```

```
Stratified Cox Model in data set expanded in two strata to allow simultaneous
assessment of covariates effect on two competing risks.
Covariates whose name is not changed have the same effect on both events.
Covariates whose name is prefixed by Main_ have effect only on the main event.
Covariates whose name is prefixed by Compet_ have effect only on the competing
event.

Stratified Cox regr. -- no ties
.....
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    Main_sex |   1.775555   .4709861     2.16   0.030     1.055707    2.986242
  Main_thick |   1.172454   .0383144     4.87   0.000     1.099714    1.250006
  Compet_sex |   1.720124   .9280661     1.01   0.315     .5974581    4.952361
Compet_thick |    1.10589   .0847623     1.31   0.189     .9516358    1.285149
------------------------------------------------------------------------------
                                                    Stratified by __000002
```

24

The option -savexpanded(*filename [, replace]*)- allows to save the dataset in the expanded format.

Data in expanded format can be used:

- to reproduce the model fitted by -stcompadj-
- to test the equality of the covariate effects on the main and competing risks
- to compare the baseline hazards for the main and competing events and test the difference under the assumption of their proportionality.

See reference [5] and help file of -stcompadj- for details.

25

Flexible Parametric Models

- Flexible parametric models have been proposed as a valid alternative to the Cox model.
- -stpm- enabled these models to be fitted by Stata[6]. Recently -stpm2[7] allowed the estimates of the confidence intervals of the fitted cumulative hazard function to be obtained too.
- When -flexible- is specified, -stcompadj- fits a flexible parametric model to the expanded data set. Then, by adding the option -ci-, the confidence intervals of the covariate-adjusted cumulative incidence are worked out from the confidence intervals of the fitted cumulative hazard function.

```
stcompadj sex=1 thick=2, compet(2) maineffect(thick) ///
                flexible ci gen(FlCI_Main FlCI_Comp) df(2)
```

26

Flexible Parametric Models

- Flexible parametric models have been proposed as a valid alternative to the Cox model.
- -stpm- enabled the fit of these models in Stata[6]. Recently -stpm2[7] extended the methodology allowing to obtain the estimates of the confidence intervals of the fitted cumulative hazard function.
- When -flexible- is specified -stcompadj- fits a flexible parametric model to the expanded data set. Then, by adding the option -ci-, the confidence intervals of the covariate-adjusted cumulative incidence are worked out from the confidence intervals of the fitted cumulative hazard function.

```
stcompadj sex=1 thick=2, compet(2) maineffect(thick) ///
                flexible ci gen(FlCI_Main FlCI_Comp) df(2)
```

27

Flexible Parametric Models

- Flexible parametric models have been proposed as a valid alternative to the Cox model.
- -stpm- enabled the fit of these models in Stata[6]. Recently -stpm2[7] extended the methodology allowing to obtain the estimates of the confidence intervals of the fitted cumulative hazard function.

The option generate(newvarname1 newvarname2) allow to give two alternative names to the variables containing the covariate-adjusted CI function for the main and competing event.

Higher and lower confidence bounds of the CI are saved in four new variables whose names are prefixed by Hi_ and Lo_.

```
stcompadj sex=1 thick=2, compet(2) maineffect(thick) ///
                flexible ci gen(FlCI_Main FlCI_Comp) df(2)
```

28

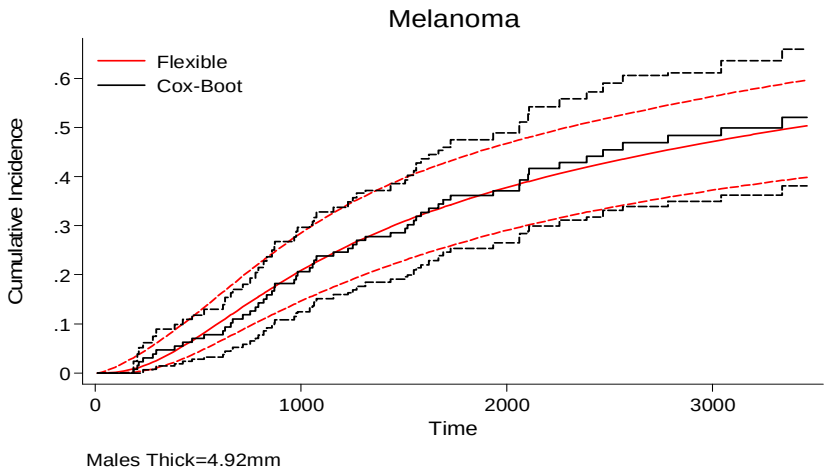
As an alternative, the option –bootci– computes the confidence intervals of the cumulative incidence (CI) function by resampling observations from the expanded dataset:

```
stcompadj sex=1 thick=2, compet(2) maineffect(thick)  ///
                    bootci gen(BootCI_Main BootCI_Comp)
```

In the next graph we compare the CI functions for the main and the competing event estimated by fitting a:

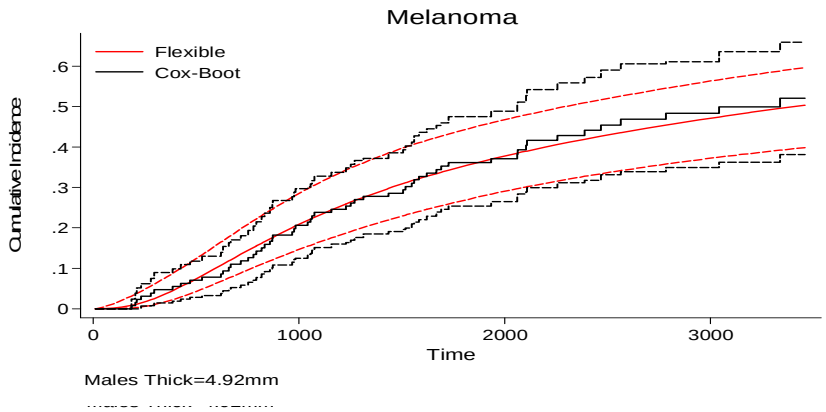
1. **Cox model**. Confidence Intervals are computed by resampling observations (1000 reps are the default);
2. **flexible parametric model**. Confidence intervals are computed from the confidence intervals of the predicted cumulative hazard function

29



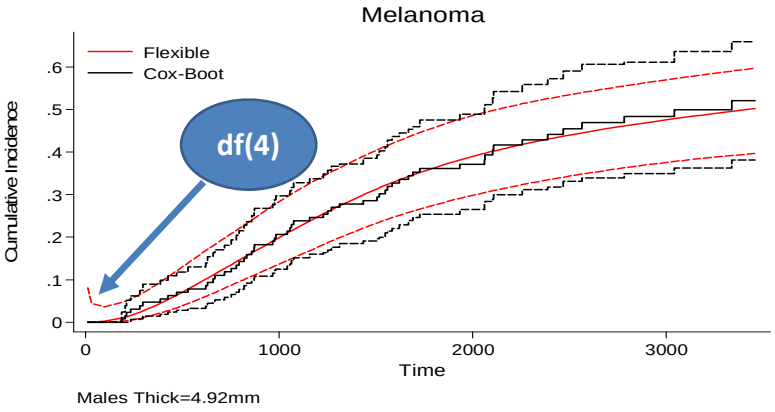
- The CI estimates (solid lines) obtained by fitting a Cox model and a flexible parametric model are in good agreement until 1500 days from the diagnosis. After, some difference can be seen.

30



- The confidence intervals (dashed lines) computed by fitting a flexible parametric model are narrower than those obtained by resampling observations. Mostly the *flexible* upper confidence bounds are lower than the *bootstrapped* confidence bounds in the right tail of the curve. I believe this is a sensible finding because parametric models usually produce Standard Error smaller than non-parametric approaches.

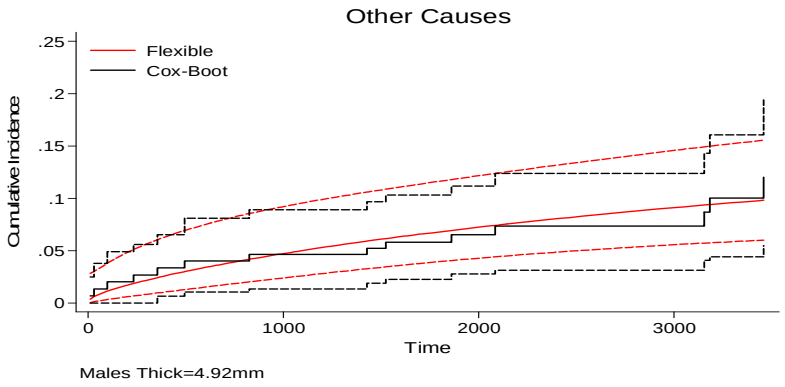
31



- In the previous example the number of `df` is set to 2

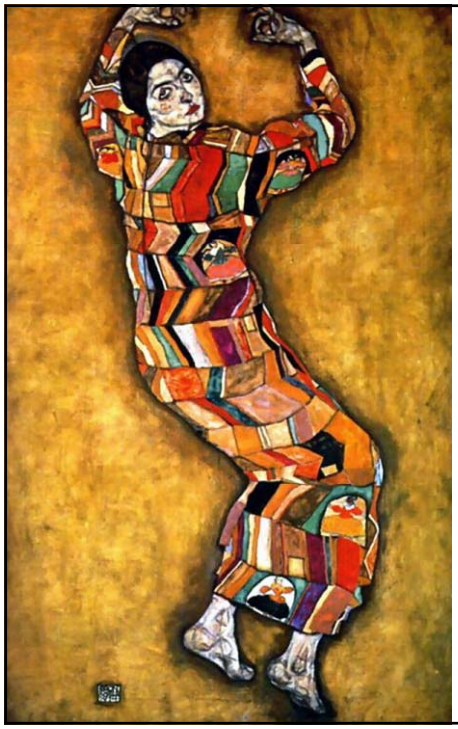
  When the events are sparse (like at the beginning of the follow-up) and more than a few dfs are specified, the upper confidence bound computed by the flexible parametric model can be large.

32



- The same considerations pertain to the CI functions of the Other Causes of death.
- Some difference between the two methods seems more evident because the flexible parametric estimates, computed at each time point, interpolate the estimates produced by the Cox approach just when an Other Cause of death occurs (14 events only).

33



Comparing covariate-adjusted cumulative incidence by two approaches

34

- Stata 11 allows competing risks regression models to be fitted.
- Fine and Gray[8] proposed a method that directly compares the cumulative incidence function by modeling the so-called hazard of the subdistribution.
- -stcrreg- is the new command that fits this model and estimates the hazard of the subdistribution ratios. Furthermore, via -stcurve-, we can obtain estimates of the CI predicted by the model.
- In the following example we will compute the CI estimates in males and females by using:
  1. the Fine and Gray regression model (-stcrreg-)
  2. the Cox model (-stcompadj-)

35

- These estimates will be compared using the cumulative incidence produced in two groups via -stcompet- (with no model assumption) to see which approach better fits the data.

```
* Cox estimate for sex
stset time,failure(cause==1) id(id)
stcompadj sex=1, compet(2)
rename CI_Main CI_Msex1
stcompadj sex=0 , compet(2)
rename CI_Main CI_Msex0
```

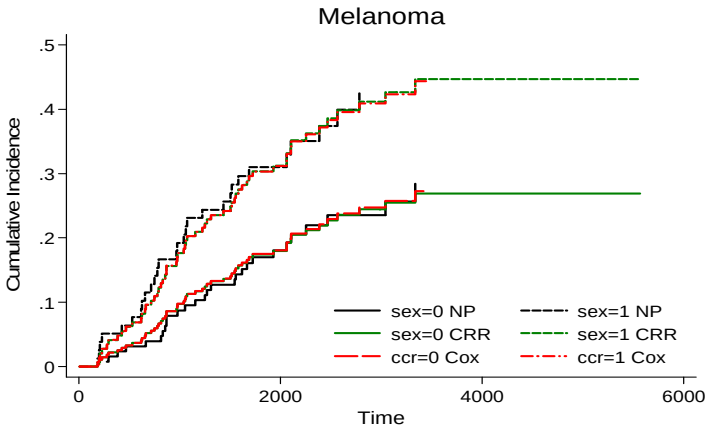
sex is modeled as having the same effect on the main and on the competing event

```
* Estimates of FG cumulative incidence
stcrreg sex, compete(cause==2)
stcurve, cif at1(sex=0) at2(sex=1) outfile(ci_strreg,replace)
```

36

```
* Non parametric CI estimates
stcompet CumInc=ci, compet1(2) by(sex)
g CI_sex0 = CumInc if sex==0 & cause==1
g CI_sex1 = CumInc if sex==1 & cause==1

merge m:m _t using ci_strreg
```

37



In this graph we see that:

- the Cox and Fine-Gray approaches provide almost exactly the same estimates
- they both reach an excellent fit to the data.

38

This is not always the case.

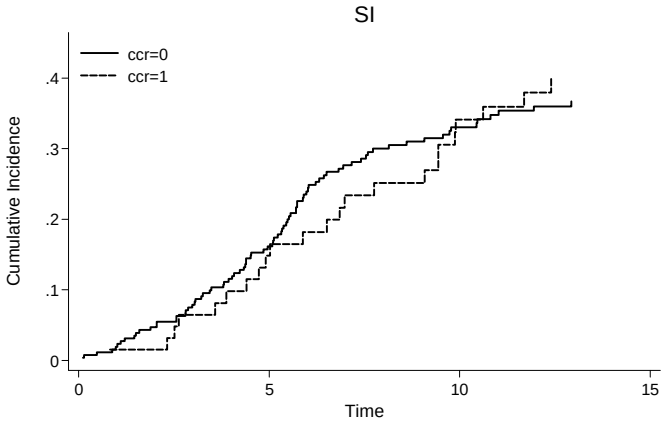
An opposed example can be found in the tutorial by Putter and coll., reproduced in the analysis of [ST] manual (p. 207-211, 226-227).

The data consists of 324 HIV infected patients. The competing events are:

1. the appearance of syncytium inducing (SI) HIV phenotype
2. AIDS

The aim of the analysis is to model the cumulative incidence in relation to the variable `ccr` that equals 1 if a deletion occurs in a receptor gene and 0 otherwise.

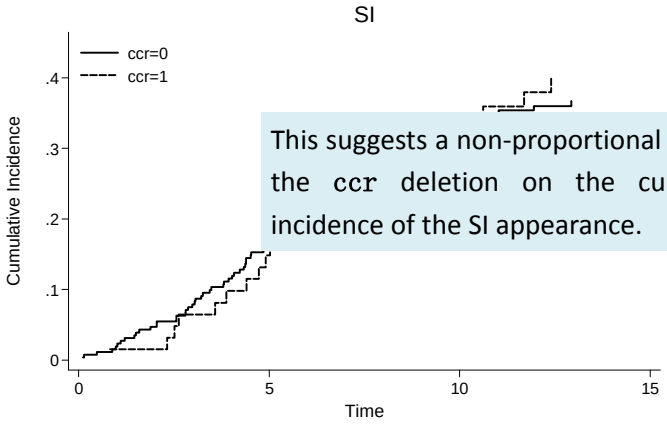
39



The cumulative incidence estimate obtained via `-stcompet-` shows that

- the subjects with `ccr` deletion have a probability of SI appearance lower than the wild-type `ccr` carriers until 10 years
- the probabilities cross and the relationship between two groups inverts after ten years have elapsed.

40



This suggests a non-proportional effect of the ccr deletion on the cumulative incidence of the SI appearance.

The cumulative incidence estimate obtained via -stcompet- shows that

- the subjects with ccr deletion have a probability of SI appearance lower than the wild-type ccr carriers until 10 years
- after, the probabilities cross and the relationship between two groups inverts.

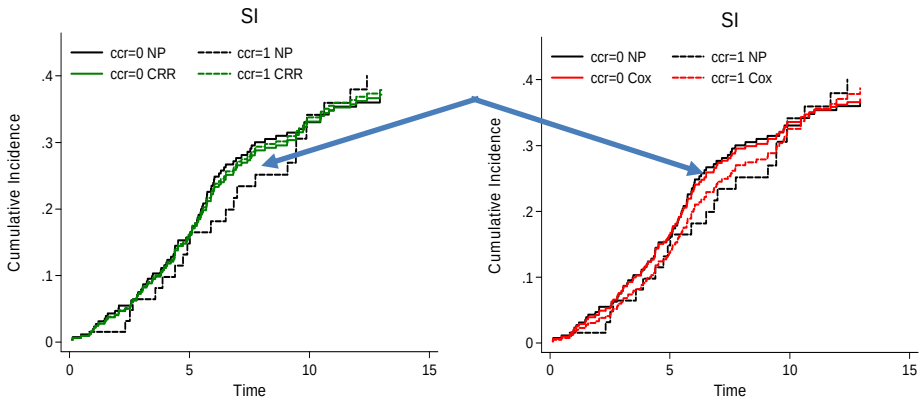
41

```
use e:\data\si,clear
stset time, failure(status==2) // SI is the event of interest

* Estimates of FG Cumulative Incidence
stcrreg ccr, compete(status==1)
stcurve, cif at1(ccr=0) at2(ccr=1) outfile(ci_strreg,replace)

* Estimates of Cox CI
stcompadj ccr=0, compet(1) maineffect(ccr) ///
       competeffect(ccr) gen(CI_Main0 CI_Compet0)

stcompadj ccr=1, compet(1) maineffect(ccr) ///
       competeffect(ccr) gen(CI_Main1 CI_Compet1)

* Non parametric CI estimates
stcompet CumInc=ci, compet1(1) by(ccr)
g CI_ccr0 = CumInc if ccr==0 & status==2
g CI_ccr1 = CumInc if ccr==1 & status==2
```

42

Non Parametric Vs. Competing Risks Regression

Non Parametric Vs. Cox Regression



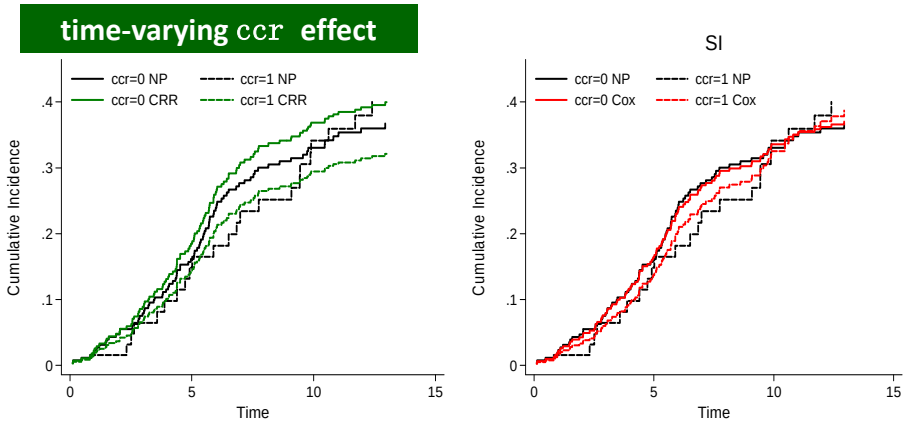
In this example the non-parametric analysis (`-stcompet-`) agrees better with the Cox (`-stcompadj-`) rather than the Fine and Gray (`-stcrreg-`) approach.

43

Non Parametric Vs. Competing Risks Regression

time-varying `ccr` effect

Non Parametric Vs. Cox Regression



The estimates produced via `-stcompadj-` fit the effect of the `ccr` variable better than the estimates predicted by `-stcrreg-`, even when a time-varying effect of `ccr` deletion is allowed in the Fine and Gray model

44

It may seem surprising that the covariate-adjusted cumulative incidence obtained by -stcompadj- (Cox model) fits the data better than the Fine and Gray model prediction. However, we should consider that

- -stcrreg- predicts the cumulative incidence function from the cumulative subhazard distribution of the main event alone;
- -stcompadj- fits two models, one for the main and one for the competing event. From each
  - it derives separate baseline hazard contributions estimates for the main and for the competing event;
  - it combines these estimates as in the non-parametric approach.
- This allows -stcompadj- to achieve a greater flexibility in estimating the cumulative incidence than the Fine and Gray approach.

45

Conclusions

- In the context of competing risks, the analysis of cause-specific hazards and the analysis of the cumulative incidence of a specific event convey different pieces of information and they both are worth to be studied.
- The estimate of the cumulative incidence function can be obtained by applying the cause-specific survival and the competing risks (Fine and Gray) regression models.
- Data augmentation technique offers several pros, particularly it consents to easily estimate the covariate-adjusted cumulative incidence according to the cause-specific approach.
- -stcompadj- is a new Stata command automating the steps required to prepare data and producing the estimate of this function at a specified level of the covariates included in the model.

46

- To this aim –stcompadj– can fit the usual Cox model or the more recent flexible parametric models, the latter allowing a straightforward estimate of the confidence intervals of the covariate-adjusted cumulative incidence.
- When the Fine and Gray model does not fit the data, a better estimate of the covariate-adjusted cumulative incidence can be achieved through the cause-specific survival approach, i.e. by –stcompadj–.
- The new command is also provided with a help file in which the user can run an example, taken from references [4] and [5], by clicking on the viewer window.
- –stcompadj– is available for download from the SSC-Archive.

47

References

1. Gooley, T.A., Leinsering, W., Crowley, J and Storer, B.E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Statistics in Medicine, 18, 695-706.
2. Pintilie, M. (2006). Competing risks. A practical perspective. Chichester: John Wiley & Sons, Inc.
3. Lunn, M., and McNeil, D. (1995). Applying Cox regression to competing risks. Biometrics, 51, 524-532.
4. Rosthøj, S., Andersen, P.K. and Abildstrom, S.Z. (2004). SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. Computer Methods and Programs in Biomedicine, 74, 69-75.
5. Putter, H., Fiocco, M., and Geskus, R.B. (2007). Tutorials in biostatistics: Competing risks and multi-state models. Statistics in Medicine, 26, 2389-2430.
6. Royston, P. (2001). Flexible parametric alternatives to the Cox model, and more. Stata Journal 1: 1–28.
7. Lambert, P.C. and Royston, P. (2009). Further development of flexible parametric models for survival analysis. Stata Journal 9: 265–290.
8. Fine, J.P and Gray, R.J. (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 94: 495-509.

48

References

1. Gooley, T.A., Leinsering, W., Crowley, J and Storer, B.E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Statistics in Medicine, 18, 695-706.
2. Pintilie, M. (2006). Competing risks. A practical perspective. Chichester: John Wiley & Sons, Inc.
3. Lunn, M., and McNeil, D. (1995). Applying Cox regression to competing risks. Biometrics, 51, 524-532.
4. Rosthøj, S., Andersen, P.K. and Abildstrom, S.Z. (2004). SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. Computer Methods and Programs in Biomedicine, 74, 69-75.
5. Putter, H., Fiocco, M., and Geskus, R.B. (2007). Tutorials in biostatistics: Competing risks and multi-state models
6. Royston, P. (2001)
1: 1–28.
7. Lambert, P.C. and
survival analysis. Stata Journal 9: 265–290.
8. Fine, J.P and Gray, R.J. (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 94: 495-509.

I thank Kerry Kanmire (from Stata tech-support) for helping me in writing the Mata code.

49

Paintings from Egon Schiele

Thanks

50