

Survival Analysis IV

Chiung-Yu Huang

April 26, 2018

ChiungYu.Huang@ucsf.edu

Competing Risks

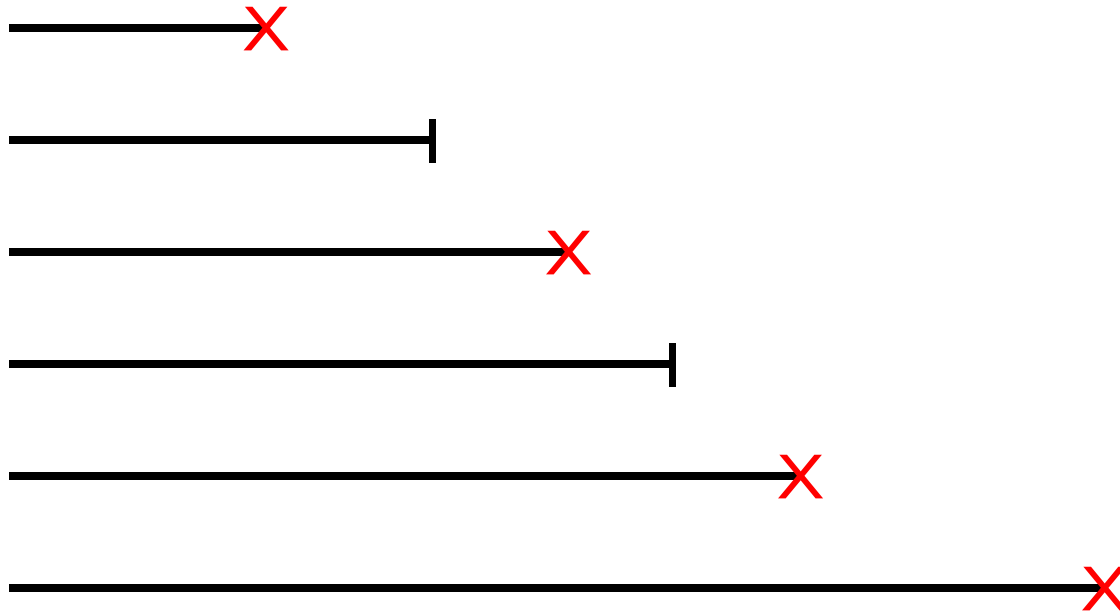
Survival Analysis

- Observed survival data:
 - Y = time to failure event or lost to follow-up, whichever occurred first
 - $\Delta = 1$ if failure event was observed
= 0 if censoring was observed
- Goal: make inference about time to failure event had subjects were followed until failure

Complication arises due to censoring

- The event time is unknown for patients who lost to follow up
- **Elimination:** extrapolate to a scenario in which censoring never happens
- This can be achieved by assuming that censoring is independent of survival time (analogous to “missing completely at random”)
- Kaplan-Meier curves estimates the survival probability in the perfect scenario where all subjects were followed until they experience the failure event

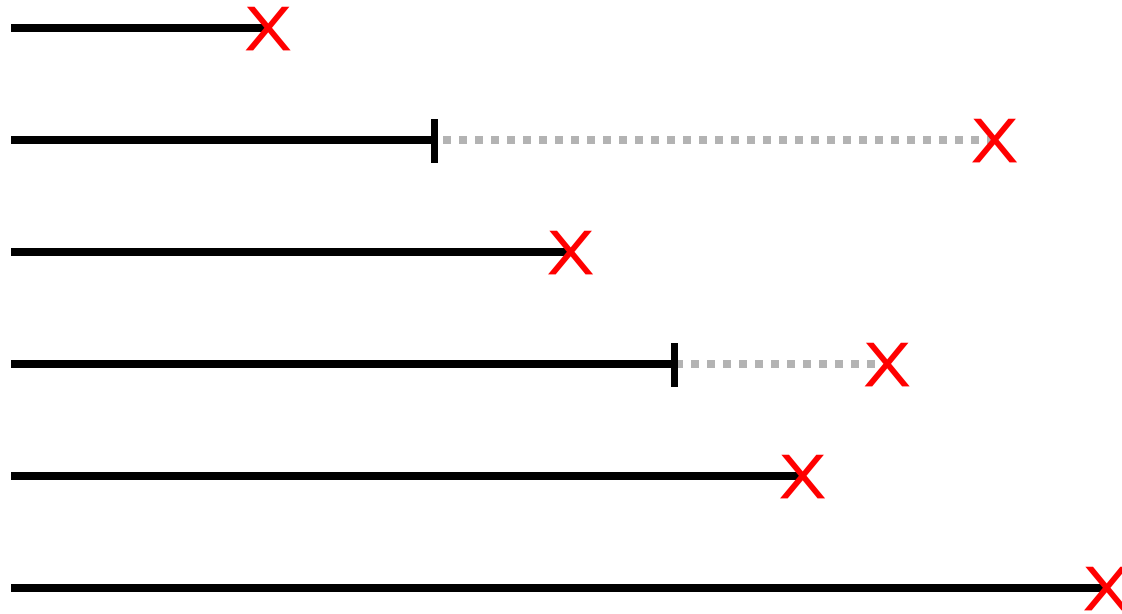
Extrapolate to eliminate incomplete follow up



X: event time

I: lost to follow up

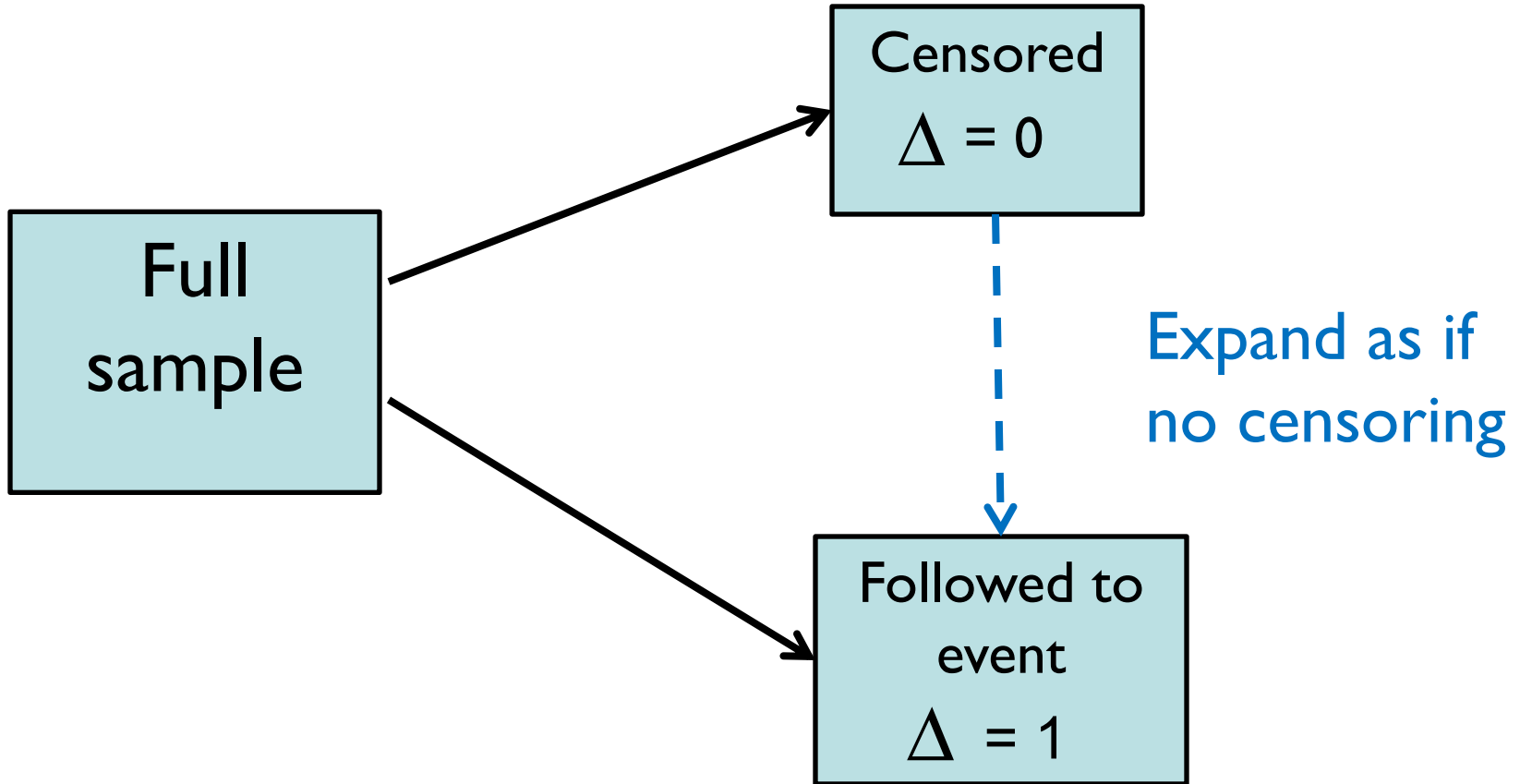
Extrapolate to eliminate incomplete follow up



X: event time

I: lost to follow up

Elimination



Competing Risks

- Competing risks analysis mostly concerned with when non-related death can occur – potentially excluding the event of interest
- Remember conventional Kaplan-Meier – subject lost to follow up
→ number of subjects “at risk” drops by one
 - Assumption: independent censoring
 - Objective: extrapolate to “eliminate” incomplete follow up
- But what about death due to “other causes”?
- Assumption and Objective both flawed
Kaplan-Meier curves are not appropriate for competing risks

Competing Risks

Definition: *Competing risks data* arise when multiple events can occur and follow-up can end due to occurrence of one or more of those types of events, precluding observation of at least one of the other event types

Observed data:

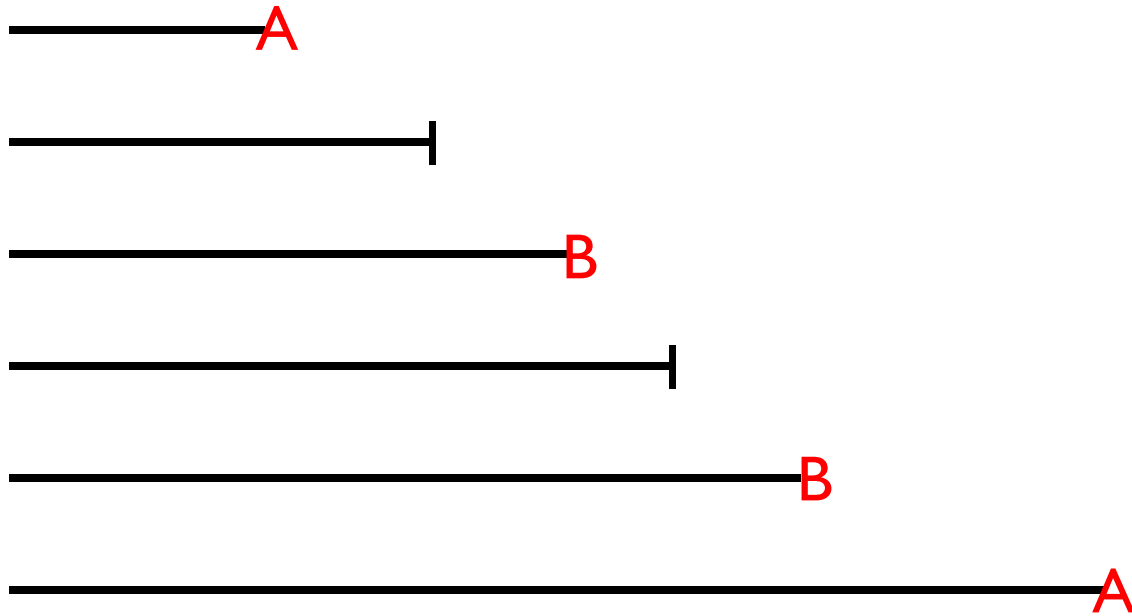
Y : time of first event of any type

$\Delta = k$ if the k -th event type occurs first

Censoring in Competing Risks

- Both **event time** and **event type** are unknown for patients who lost to follow-up
- **Elimination:** extrapolate to a scenario in which censoring never happens
- This can be achieved by assuming that censoring is independent of event time and event type

Competing risks data

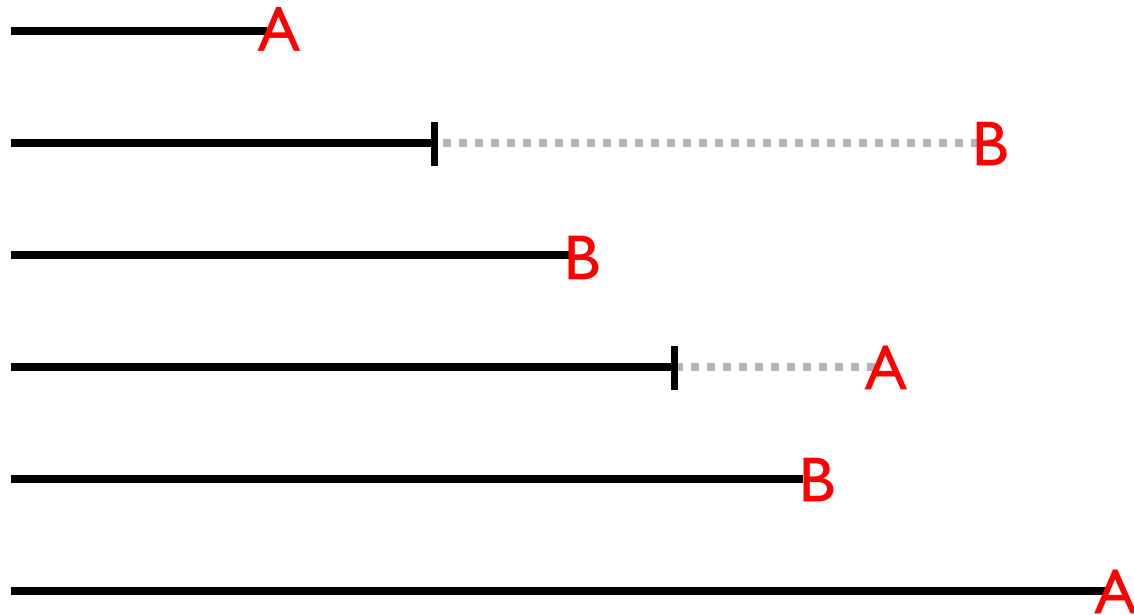


A: fracture;

B: death without fracture;

I: lost to follow up

Extrapolate to eliminate incomplete follow up



A: fracture;

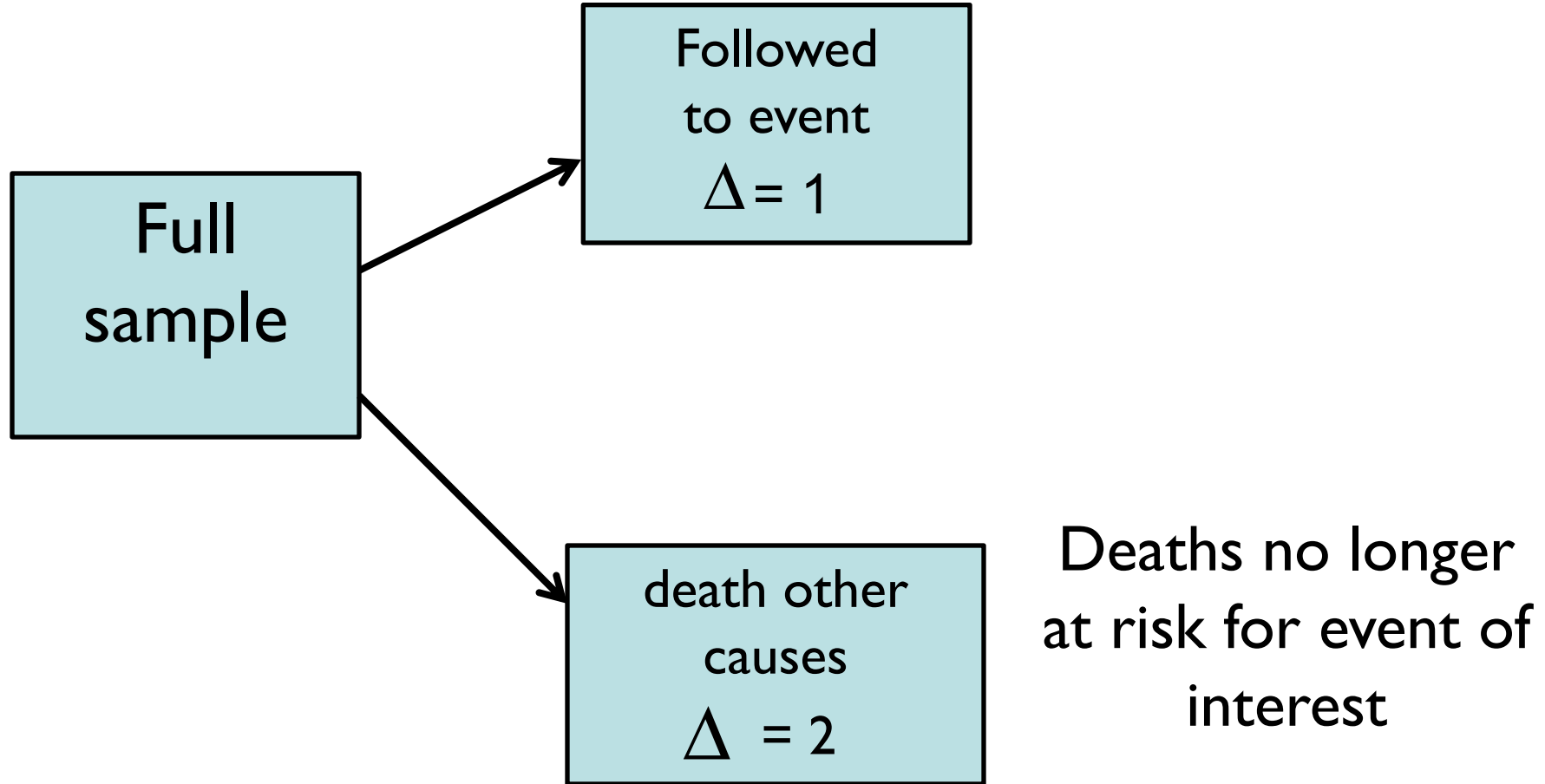
B: death without fracture;

I: lost to follow up

Competing Risks

- **Elimination:** extrapolate to a scenario in which a type of event never happens [e.g., “loss to follow up” or “incomplete follow up”]
- **Accommodation:** acknowledge and allow for the competing risks – treat the competing risk as excluding the patient from having the event of interest [e.g., death]

Accommodation



MrOS study

- Study predictors of bone fracture and low bone mineral density (BMD)
- 5993 men over 65 with average 5 year follow-up
- 531 participants fracture, 657 deaths prior to fracture, 4805 alive without fracture

MrOS study

- How well does baseline BMD predict fracture?
- Two sources of incomplete follow-up:
 - 1) end of observation (loss to follow-up / short observation times due to staggered entry)
 - 2) death

MrOS data

```
clear
use mros.dta
. list years status bmd3 weight in 1/7
```

	years	status	bmd3	weight
1.	6.518823	0	3	87.8
2.	6.499658	0	2	99.4
3.	6.49692	0	2	83.1
4.	6.49692	0	3	97.4
5.	3.036277	2	3	82.9
6.	6.494182	0	3	76.3
7.	4.50924	1	1	62.8

status

= 0 if censored

= 1 if fracture

= 2 if deaths prior to fracture

Cause-Specific Hazard Function

Definition: The *cause-specific hazard function* for event type, k , $h_k(t)$, is the short-term rate at which subjects experience the onset of the k -th event among those who have not yet experienced the event of interest (e.g., fracture) or a competing event (e.g., death) prior to t .

observed k -th events at time t

at risk of k -th event at time t

Cause-Specific Hazard Function

observed k-th events at time t

at risk of k-th event at time t

at risk of k-th event at time t = # alive
without fracture immediately prior to time t

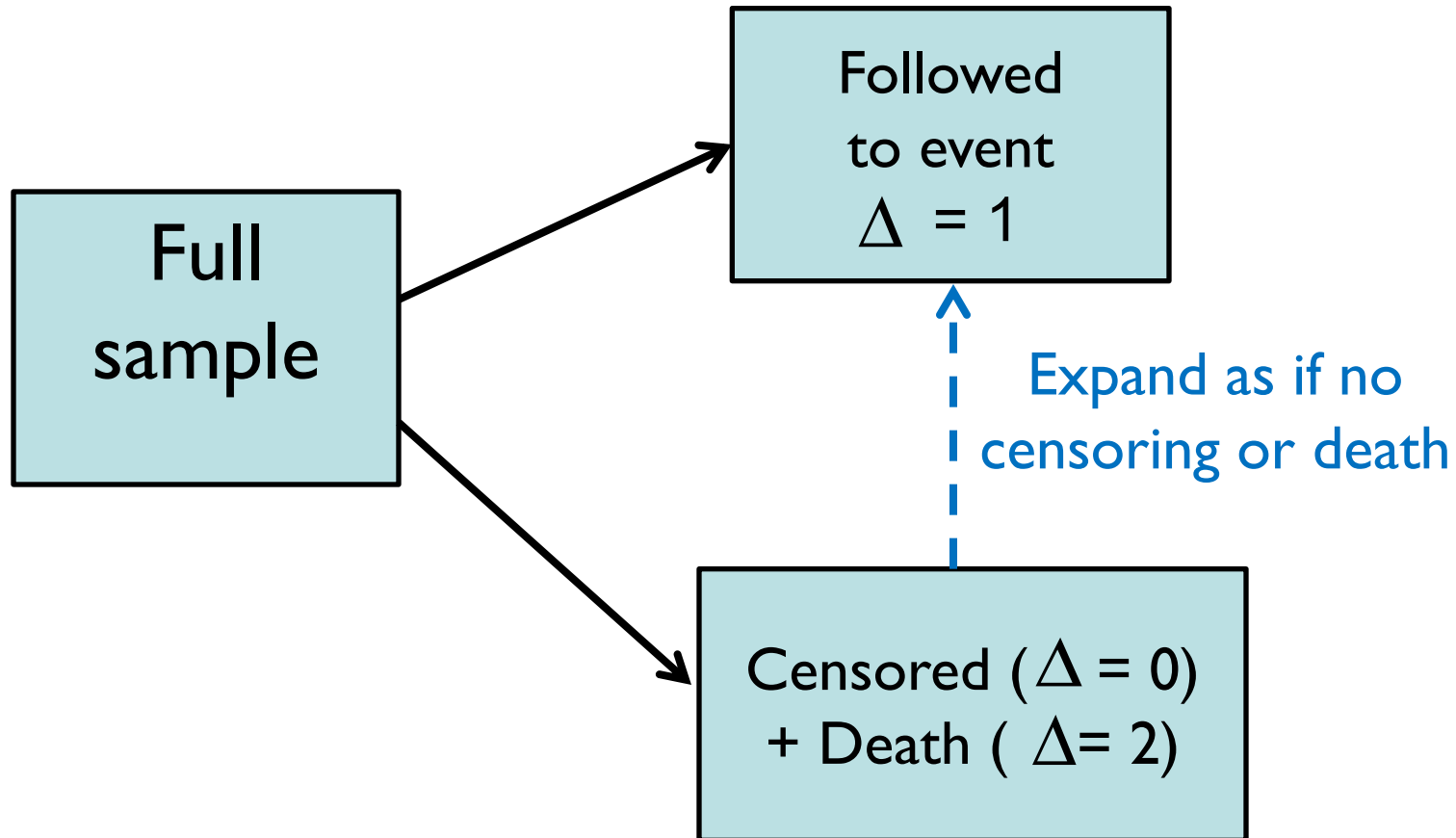
To estimate cause-specific hazard simply set up
data as ordinary survival with k-th failure as the
only type – all competing causes (even death)
treated as censored

Cause-Specific Hazard Function

Counterintuitively – **cause-specific hazard function** does not make a distinction between censoring and death...

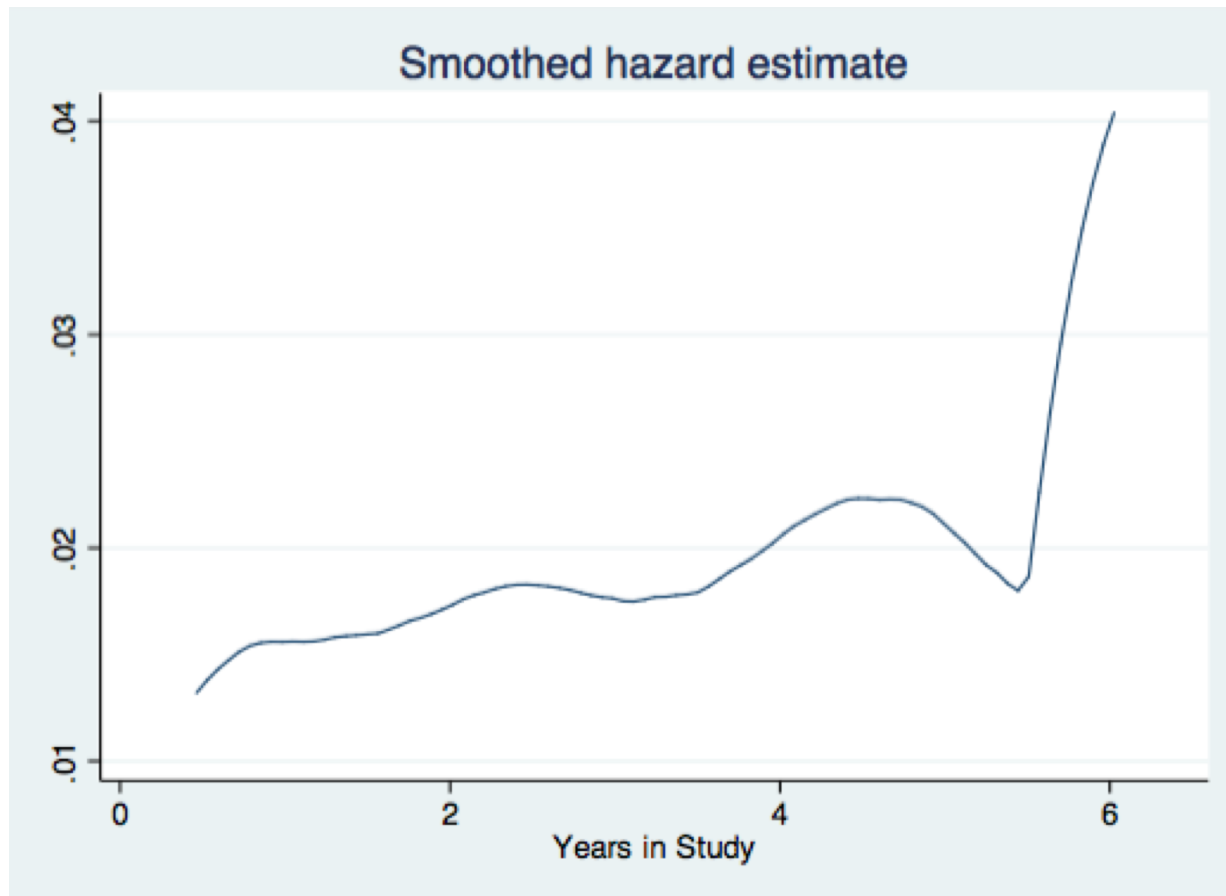
contrast with *cumulative incidence function*...

Elimination



MrOS cause-specific hazard

```
use mros.dta  
stset years, failure(status==1) // fracture is the event  
sts graph if _t < 6.5, hazard xtitle(Years in Study)
```



Cox model for cause-specific hazard function

- Model cause-specific hazard function using Cox proportional hazard formulation
- Let $h_k(t)$ be the k -th cause-specific hazard, then...
- $h_k(t|x) = h_{0k}(t) \exp(\beta_{1k}x_1 + \dots + \beta_{pk}x_p)$
- Works same as original Cox model except HRs apply to the k -th cause specific hazard

MrOS Cox model for cause-specific hazard function

```
clear
use mros.dta
stset years, failure(status==1)
stcox i.bmd3 weight, nolog // bmd is defined in 3 quantiles
```

No. of subjects =	5993	Number of obs =	5993
No. of failures =	531		
Time at risk =	30483.46339		
		LR chi2(3) =	121.22
Log likelihood =	-4442.2904	Prob > chi2 =	0.0000

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
bmd3						
2	.4193745	.0447953	-8.14	0.000	.3401585	.5170384
3	.3290229	.0396476	-9.23	0.000	.2598098	.4166743
weight	1.004146	.00362	1.15	0.251	.997076	1.011266

Cause-specific hazard regression for death without fracture

```
stset years, failure(status==2)  
stcox i.bmd3 weight, nolog // bmd is defined in 3 quantiles
```

No. of subjects =	5993	Number of obs =	5993
No. of failures =	657		
Time at risk =	30483.46339		
Log likelihood =	-5525.1007	LR chi2(3) =	27.64
		Prob > chi2 =	0.0000

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
bmd3						
2	.9477798	.0894883	-0.57	0.570	.7876585	1.140452
3	.8260441	.0861232	-1.83	0.067	.6733752	1.013326
weight	.9871025	.0033244	-3.85	0.000	.9806082	.9936398

Cumulative Incidence Function

Definition: The *cumulative incidence function* for cause k at time t , $F_k(t)$, is the proportion who have developed the k -th event prior to t

The definition above makes adjustment to “# at risk” for censoring (elimination), but not for other competing risks, i.e., for death in the MrOS dataset (accommodation)

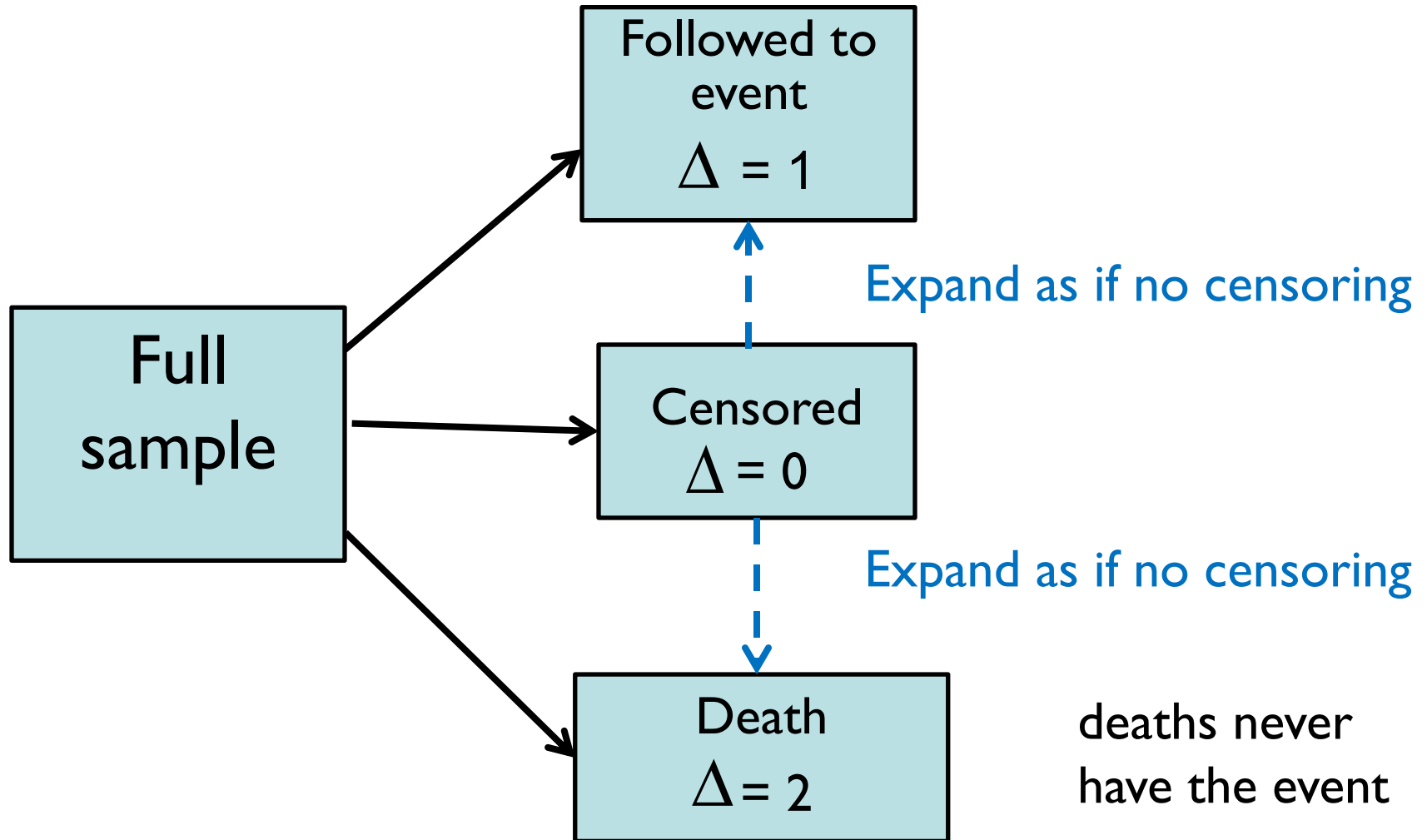
Cumulative Incidence Function

The cumulative incidence function for the k -th event at time t , $F_k(t)$, is the probability that a subject experienced k -th event at or before time t

$$F_k(t) = F_k(t-1) + \tilde{S}(t-1)h_k(t)$$

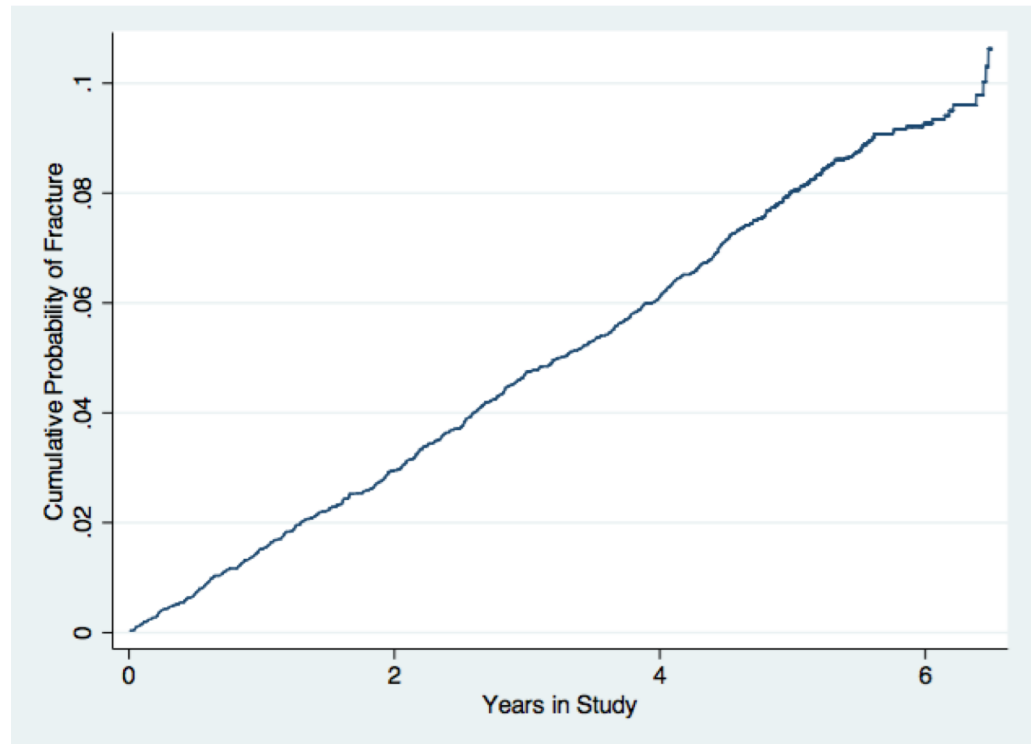
where $\tilde{S}(t-1)$ is the probability of being free of events at time $t-1$, i.e. just prior to time t . (In the MrOS data = chance of being alive and unfractured)

Accommodation for competing events plus elimination for censoring



MrOS cumulative incidence function

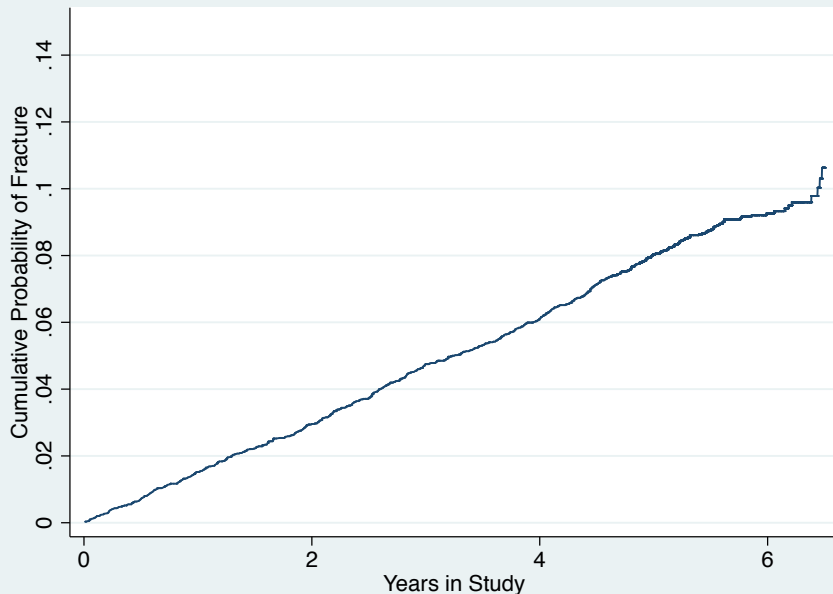
```
clear
use mros.dta
stset years, failure(status==1) *This specifies event of interest
stcrreg, compete(status==2) *This specifies competing event
predict cif_fracture, basecif
twoway (line cif_fracture _t if _t < 6.5, sort), ///
ytitle(Cumulative Probability of Fracture) ///
xtitle(Years in Study)
```



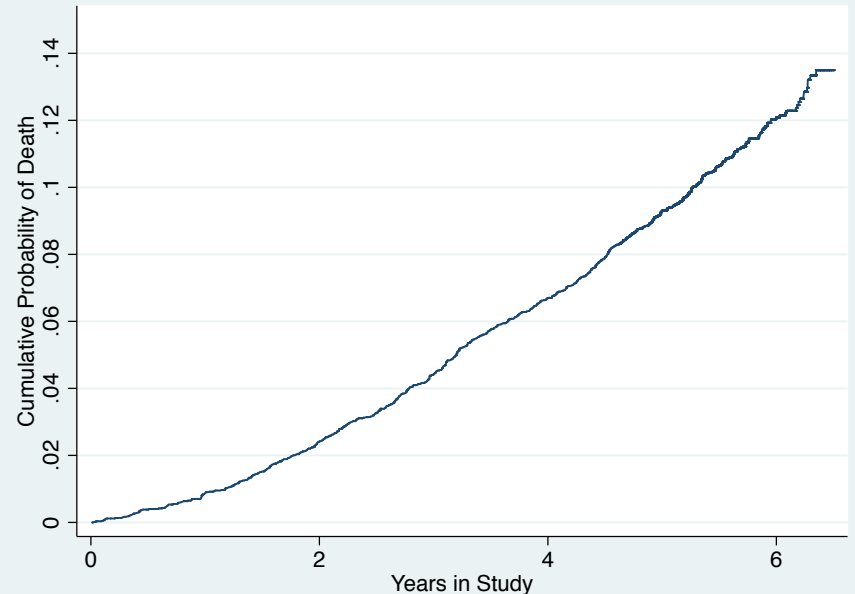
CIF for both competing events

```
stset years, failure(status==2)
stcrreg, compete(status==1)
predict cif_death, basecif
twoway (line cif_death _t if _t<6.5, sort), ///
ytitle(Cumulative Probability of Death Without Fracture) ///
xtitle(Years in Study) ylabel(0(0.02).15)
```

CIF of fracture



CIF of death without fracture



Fine-Gray model for cumulative incidence function

- What if BMD is correlated with factors that make death more likely?
- Low BMD might lead to fewer fractures after 2 years because men with low BMD are more likely to die before fracture.
- Solution: compare cumulative incidence function rather than cause-specific hazard.

Fine-Gray model for cumulative incidence function

- Model cumulative incidence function, adapting Cox-type proportional hazard formulation
- Let $f_k(t)$ be the k -th “hazard” where the cohort members who succumb to the competing risk are retained in the denominator of their rate (c.f., death)
- $f_k(t|x) = f_{0k}(t) \exp(\beta_{1k}x_1 + \dots + \beta_{pk}x_p)$
- Note: Maintaining deaths in the risk set acknowledges someone lost to competing risk will not develop event of interest so no extrapolation required

MrOS Fine-Gray model

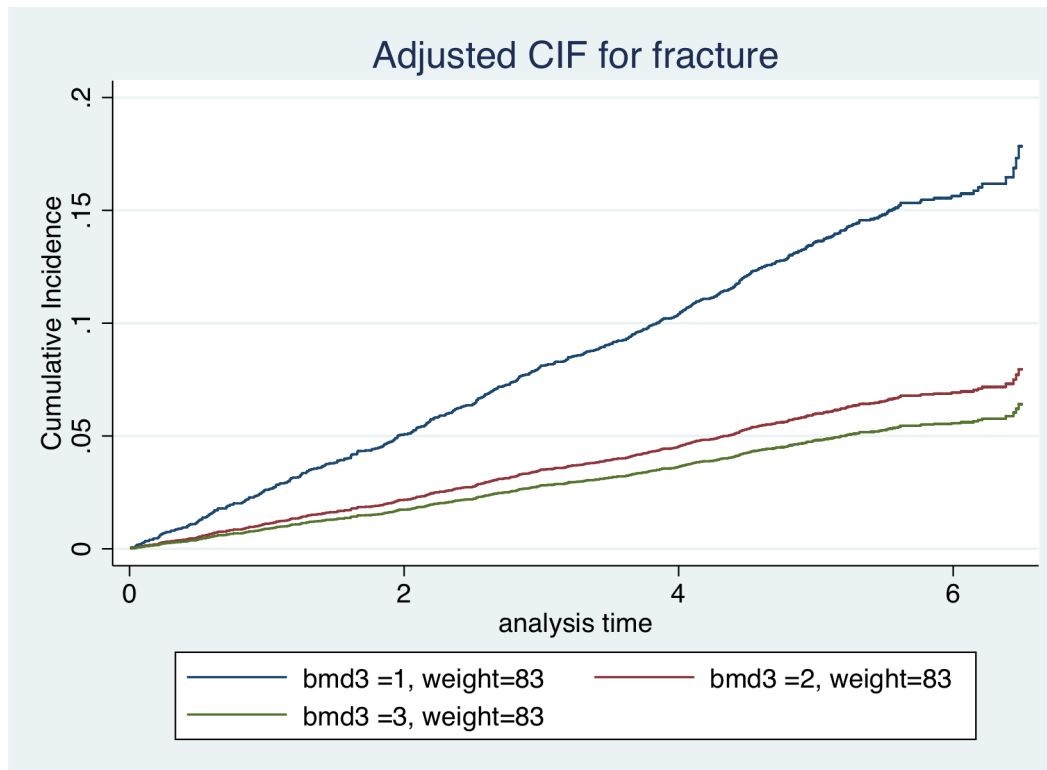
```
clear
use mros.dta
stset years, failure(status==1)
stcrreg i.bmd3 weight, compete(status==2) // bmd in 3 quantiles
```

Competing-risks regression	No. of obs	=	5993
	No. of subjects	=	5993
Failure event : status == 1	No. failed	=	531
Competing event: status == 2	No. competing	=	657
	No. censored	=	4805
	Wald chi2(3)	=	119.64
Log pseudolikelihood = -4472.9261	Prob > chi2	=	0.0000

_t	SHR	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
bmd3						
2	.4219663	.0443983	-8.20	0.000	.3433337	.5186079
3	.3369128	.03982	-9.20	0.000	.2672472	.4247387
weight	1.004669	.0036759	1.27	0.203	.9974896	1.011899

Adjusted CIF curves

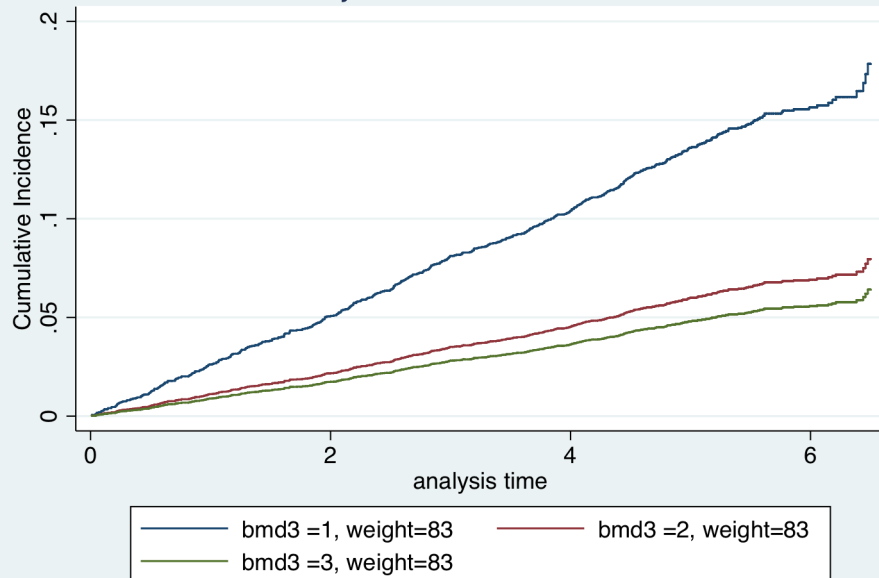
```
stcurve, cif at1(bmd3 =1, weight=83) at2(bmd3 =2, weight=83) ///  
at3(bmd3 =3, weight=83) range(0 6.5) ///  
title("Adjusted CIF for fracture")
```



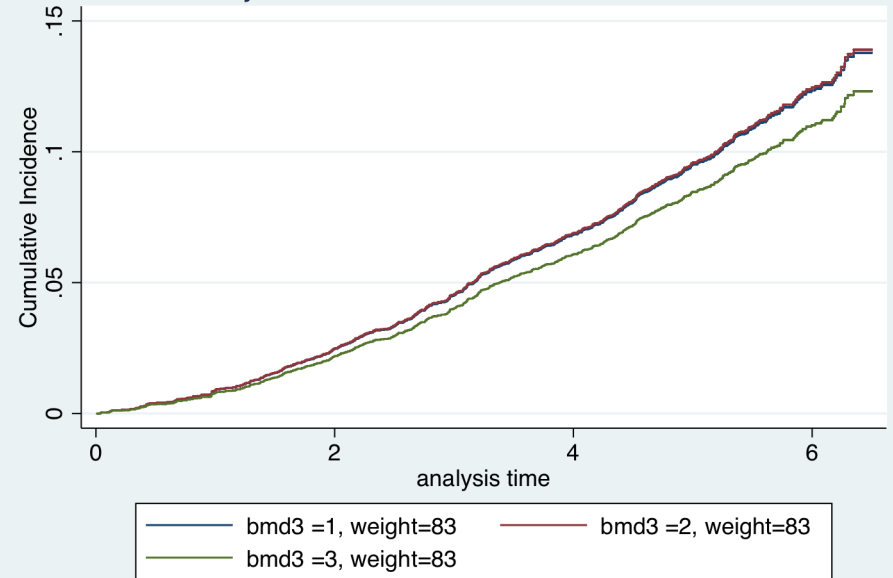
Adjusted CIF curves for death without fracture

```
stset years, failure(status==2)
stcrreg i.bmd3 weight, compete(status==1)
stcurve, cif at1(bmd3 =1, weight=83) at2(bmd3 =2, weight=83) ///
at3(bmd3 =3, weight=83) range(0 6.5) ///
title("Adjusted CIF for death without fracture")
```

Adjusted CIF for fracture



Adjusted CIF for death without fracture



MrOS model comparisons

Cause-specific Cox model

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
bmd3						
2	.4193745	.0447953	-8.14	0.000	.3401585	.5170384
3	.3290229	.0396476	-9.23	0.000	.2598098	.4166743
weight	1.004146	.00362	1.15	0.251	.997076	1.011266

Fine-Gray model

_t	SHR	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
bmd3						
2	.4219663	.0443983	-8.20	0.000	.3433337	.5186079
3	.3369128	.03982	-9.20	0.000	.2672472	.4247387
weight	1.004669	.0036759	1.27	0.203	.9974896	1.011899

MrOS model comparisons

- MrOS cause-specific hazards and Fine-Gray model results are very similar
- Makes sense: Methods differ based on whether people who die are retained in the risk set – but the risk of death is not large.
- In other cases, covariate effects may of different directions in these two approaches.
- Neither approach is incorrect; all depends on desired interpretation

Competing Risks Summary

- “Deaths” occur due to other causes before event of interest
- Elimination vs. Accommodation
- Cause-specific hazard; cause-specific Cox model (`stcox`); elimination only
- Cumulative incidence function; Fine-Gray model (`stcrreg`); elimination for censoring and accommodation for competing death (by other causes)

Other Survival Topics

- **Interval Censoring**

- Conventional survival data is right-censored only
- Often only know event happens in an interval (e.g., between scheduled hospital visits).
- Regular intervals can use pooled logistic regression (e.g., all patients visit at time 0, 12 months, 24 months and 36 months)
- Non-regular intervals need non-parametric Kaplan-Meier-Turnbull or parametric modeling (i.e. with a model for baseline hazard or survival distribution)
- Use `intcens` in STATA from SSC for parametric models
type `ssc install intcens` in STATA then look at:
`help intcens`

Other Survival Topics

- **Left-Truncation**

- PBC used time from enrolment for time 0, but makes more sense to use diagnosis time
- Why LT a problem? subjects with early death less likely to be enrolled – if not accounted for will underestimate early deaths
- `stset years_since_diag, failure(status)`
`enter(disease_dur)`
`disease_dur = truncation times (time from Dx to enrollment)`
- Survival function drops faster early on after accounting for left-truncation
- See section 6.6.6 in VGSM

Other Survival Topics

- **Clustered data**

- Dr. McCulloch's lectures have a strong focus on the analysis of clustered data and you will see many examples (but not survival)
- Clustered survival data consist of e.g., multiple subjects clustered by center, or multiple events (multiple failure times) on a subject (c.f. Dr. McCulloch's lectures to come)
- use `shared(cluster_id)`, `cluster(id)` or `vce(cluster id)` in Stata with `stcox`
- The key is to appropriately set up the data for the type of clustering – ordered vs. non-ordered events, single or multiple event types etc.
- See

<http://www.stata.com/support/faqs/statistics/multiple-failure-time-data>

Finally...

- Dr. McCulloch's first lecture on Repeated Measures coming Tuesday
- Check syllabus – all solutions (labs + homeworks) for my lectures should be posted by weekend – let me know if not
- Final project assignments will be in place today