COMPETING RISKS

Dr. Aric LaBarr

Institute for Advanced Analytics

INTRODUCTION THROUGH EXAMPLES

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

WHAT IS THE MORTALITY RATE?

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

WHAT IS THE MORTALITY RATE?

DOES 7% FEEL RIGHT?

Customer Example

- Observe customers over the past year to try and analyze **voluntary** churn.
- Of the 1000 customers in the data set, 240 left voluntarily, while 60 left involuntarily.

WHAT IS THE CUSTOMER CHURN RATE?

Customer Example

- Observe customers over the past year to try and analyze **voluntary** churn.
- Of the 1000 customers in the data set, 240 left voluntarily, while 60 left involuntarily.

WHAT IS THE CUSTOMER CHURN RATE?

DOES 30% FEEL RIGHT?

Fixed vs. Random Censoring

- **Fixed censoring** – censoring only occurs at the end of the study ($C_i = c$ is known in advance).
  - Recidivism data: Not arrested in 52 weeks is censored by design because that is when study ended.

- **Random censoring** – C_i may vary between subjects for reasons beyond the investigator's control.
  - Recidivism data: No arrest within first 30 weeks, but lose contact with subject for whatever reason.
  - Recidivism data: Study done only for one year, but people can have delayed entry into the study (as they were released).

COMPETING RISKS

Multiple Event Types

- All of the models used so far have been for studying the time until **one** event occurs.
- All of the models used so far can be extended to studying **multiple** events or **multiple types** of events.

Competing Risks

- Examples:
  - Death from cancer in medical study vs. other causes of death.
  - Leaving job due to retirement, injury, or being fired.
  - Pump failure due to jamming, flooding, motor failure, or surge.
- In all of the above cases there are multiple, **mutually exclusive** causes of failure.
- These are examples of a **competing risks** problem, where each subject can experience only one of several possible events.

Independence Again…

- Assume T_i and C_i are independent – subjects censored at time *t* were randomly selected to be censored from all subjects still in the risk set at *t*.
- **IF** this is true, then fixed vs. random censoring is mathematically equivalent.
- What does independence "mean" here?
  - In competing risks, independence implies that a censored observation and an uncensored observation have the same **same risk of the event**, **regardless of the reason for censoring**.

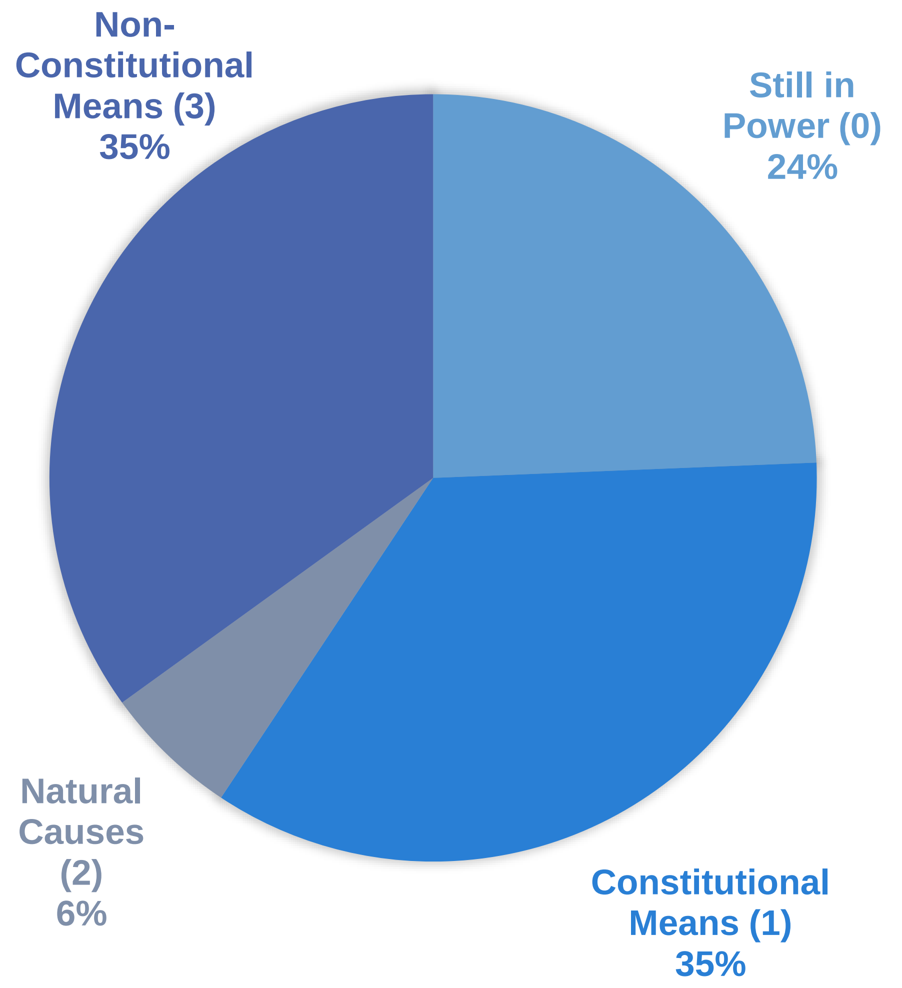
Independence Again…

- Example:
  - By treating other failure types as censored, we're essentially implying once a pump fails due to jamming, we still don't know when it would fail due to flooding – we assume that the event types are independent.
- **NO TEST FOR THIS!**
- Decide independent or not based on context of problem.
- In other words, are observations with a high risk of one event equally likely to experience the other events?

ESTIMATION

World Leaders Data Set

- Compiled by Bienen and van de Walle in 1991.
- Primary leaders of all countries between 1960 and 1987.
- Number of years the leader was in power and the manner they lost power.



World Leaders Data Set

- Manner – how the leader reached power (0: constitutional, 1: non-constitutional)
- Start – year of entry to power
- Military – background of leader (1: military, 0: civilian)
- Age – age at time of entry
- Conflict – level of ethnic conflict (1: medium/high, 0:low)
- LogInc – log of GNP per capita
- Growth – avg. annual growth rate of GNP
- Pop – population in millions
- Land – land area in 1000 km^2
- Literacy – literacy rate (unknown year)
- Region – 0: Middle East, 1: Africa, 2: Asia, 3: Latin America

Review

- Two major functions in survival analysis:
- Survival Function – probability of surviving beyond time *t:*

$$S(t) = P(T > t) = 1 - F(t)$$

- Hazard Function – conditional failure rate in an interval:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t}$$

Cause-Specific Hazard Function

- When there are multiple event types, the hazard function contains two variables – *T* and *J*.
- The cause/type specific hazard function is as follows:

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{P(t \le T_i < t + \Delta t, J_i = j \mid T_i \ge t)}{\Delta t}$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

Cause-Specific Hazard Function

- When there are multiple event types, the hazard function contains two variables – *T* and *J*.
- The cause/type specific hazard function is as follows:

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_i < t + \Delta t, \boxed{J_i = j}\, | T_i \geq t)}{\Delta t}$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

Cumulative Incidence Function

- The **cumulative incidence function** (CIF) is the *unconditional* probability that event type *k* occurs *by* time *t*:

$$F_k(t) = P(T \leq t, K = k)$$

- The probability of **any** event by time *t* is just the sum of the individual CIF's:

$$F(t) = \sum_k F_k(t)$$

Overall Survival

- Survival Function – probability of surviving beyond time *t:*

$$S(t) = P(T > t) = 1 - F(t) = 1 - \sum_k F_k(t)$$

- The overall survival function is still unconditional, since survival means surviving **all** of the risks, so there's no such thing as a cause/type specific survival.

Estimating the CIF

- We can estimate the CIF's nonparametrically using the nonparametric estimates of the survival and hazard functions:

$$\hat{F}_k(t) = \sum_{t_m \leq t} \hat{h}_k(t_m)\hat{S}(t_{m-1})$$

Estimating the CIF’s – SAS

```
proc lifetest data=Survival.leaders plots=cif(CL)
              conftype=loglog error=delta outcif=cif_est;
   time years*lost(0) / failcode=1 2 3;
   format lost lost.;
run;

proc sgplot data=cif_est;
   series x=years y=CIF / group=failcode;
   format failcode lost.;
run;
quit;
```

Estimating the CIF’s – SAS

The LIFETEST Procedure

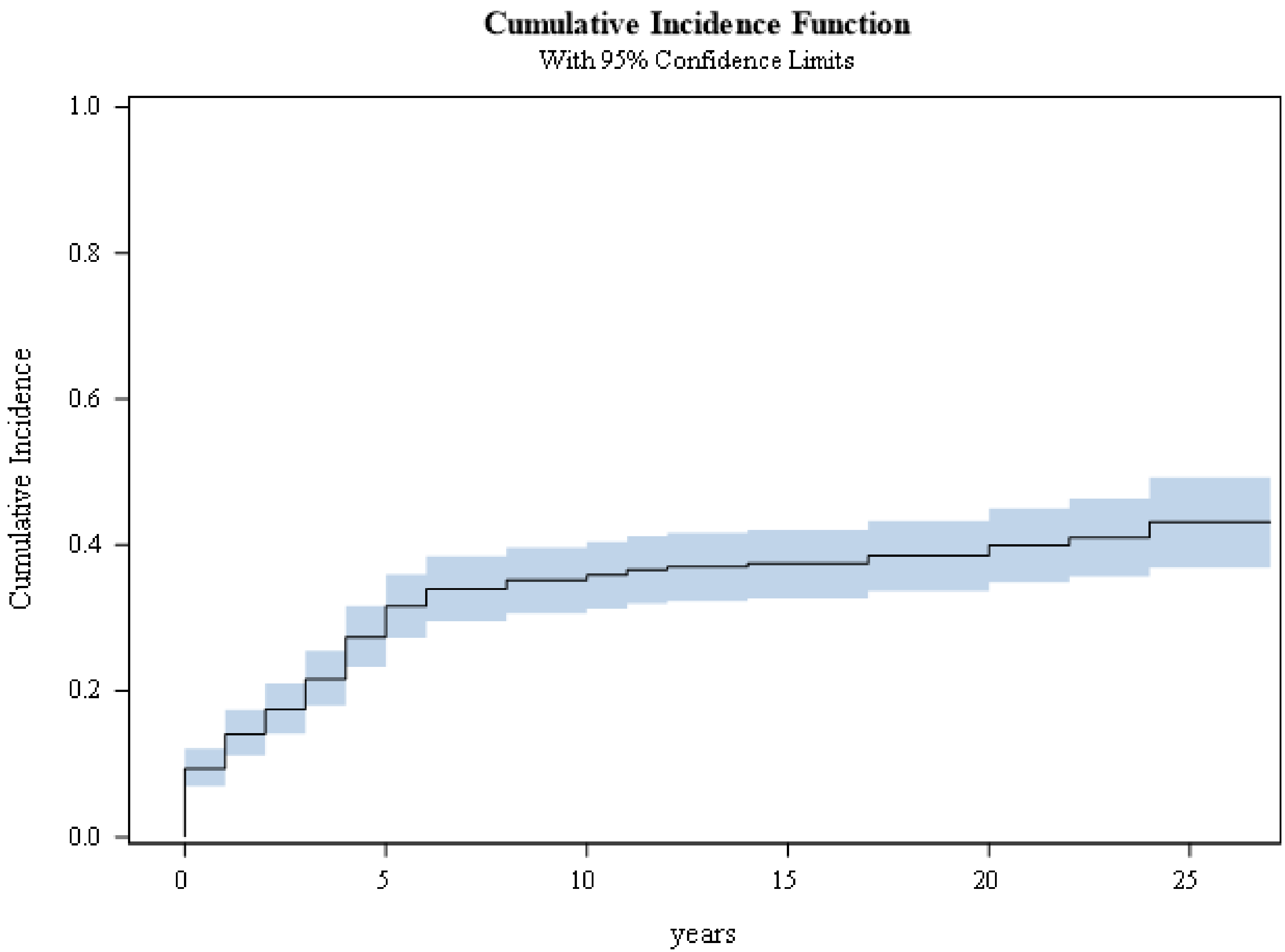
Failed Event: lost=Constitutional

Summary of Failure Outcomes			
Failed Events	Competing Events	Censored	Total
165	192	115	472

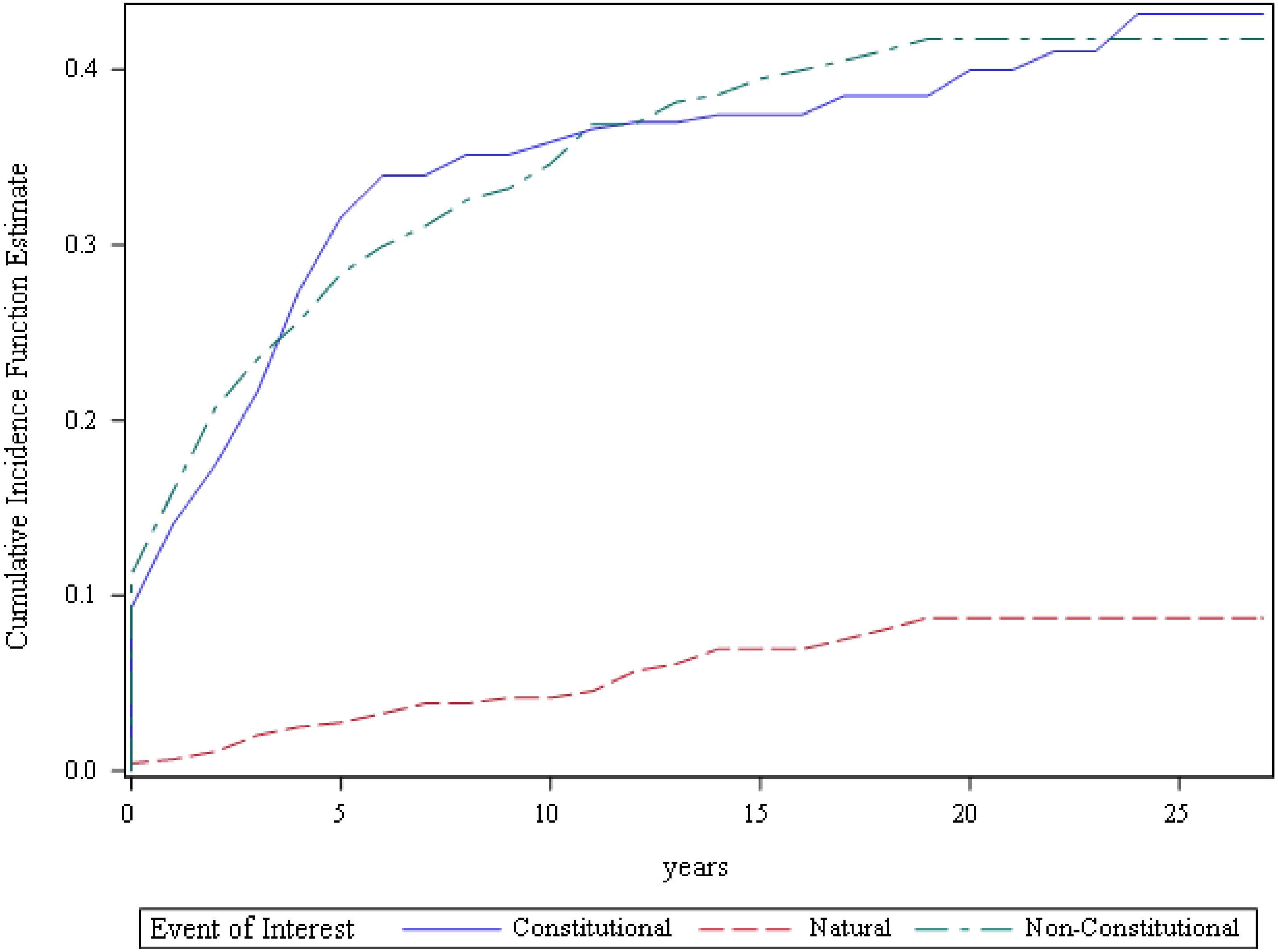
Cumulative Incidence Function Estimates				
years	Cumulative Incidence	Standard Error	95% Confidence Interval	
0	0	0	.	.
0	0.0932	0.0134	0.0691	0.1215
1	0.1407	0.0161	0.1110	0.1739
2	0.1744	0.0176	0.1414	0.2103
3	0.2161	0.0193	0.1796	0.2550
4	0.2738	0.0211	0.2332	0.3159

⋮

Estimating the CIF's – SAS



Estimating the CIF’s – SAS



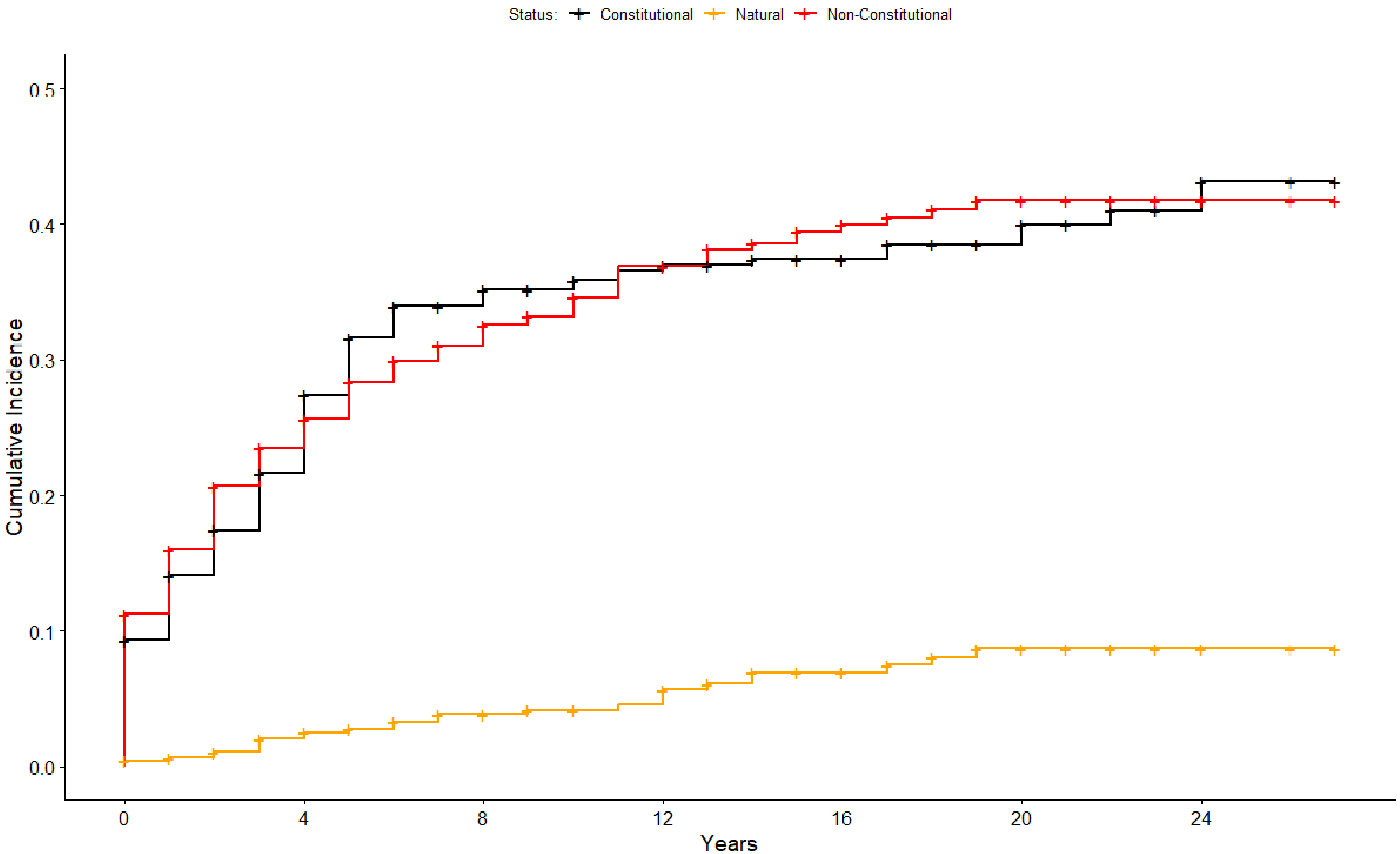
Estimating the CIF’s – R

```
lcon_data <- finegray(Surv(years, lost) ~ .,
                      data = leaders, etype = "Constitutional")
lnat_data <- finegray(Surv(years, lost) ~ .,
                      data = leaders, etype = "Natural")
lnon_data <- finegray(Surv(years, lost) ~ .,
                      data = leaders, etype = "Non-Constitutional")

lcon <- survfit(Surv(fgstart, fgstop, fgstatus) ~ 1,
                data = lcon_data, weight = fgwt)
lnat <- survfit(Surv(fgstart, fgstop, fgstatus) ~ 1,
                data = lnat_data, weight = fgwt)
lnon <- survfit(Surv(fgstart, fgstop, fgstatus) ~ 1,
                data = lnon_data, weight = fgwt)
```

Estimating the CIF's – R

```
leadlist <- list(constitutional = lcon, natural = lnat,
                 nonconstitutional = lnon)

ggsurvplot(leadlist, combine = TRUE, fun = "event",
           conf.int = FALSE, break.x.by = 4, ylim = c(0, 0.5),
           legend.title = "Status:",
           legend.labs = c("Constitutional", "Natural",
                           "Non-Constitutional"),
           break.y.by = 0.1, xlab = "Years",
           ylab = "Cumulative Incidence",
           palette = c("black", "orange", "red"))
```

Estimating the CIF's – R



CAUSE-SPECIFIC HAZARD MODEL

Modeling Type-Specific Events

- Type-Specific events can be modeled with both proportional hazard models …

$$\log h_k(t) = \log h_{0,k}(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

- … and accelerated failure time (AFT) models :

$$\log T_{i,k} = \beta_0 + \beta_1 x_{i,1} + \cdots + \sigma e_i$$

Cox Regression Competing Risks

- Typical modeling approach for competing risks is to use separate Cox regression models for **each** cause, treating all other events as censored.
- Essentially, modeling the effects of predictors on the cause-specific hazard:

$$\log h_k(t) = \log h_{0,k}(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

Cox Competing Risks – SAS

```
proc phreg data=Survival.Leaders;
   class region;
   model years*lost(0,1,3) = manner start military age
                             conflict loginc growth pop
                             land literacy region /
                             ties=efron;
run;
```

Cox Competing Risks – SAS

The PHREG Procedure

Model Information	
Data Set	SURVIVAL.LEADERS
Dependent Variable	years
Censoring Variable	lost
Censoring Value(s)	0 1 3
Ties Handling	EFRON

Number of Observations Read	472
Number of Observations Used	438

Class Level Information				
Class	**Value**	**Design Variables**		
region	**0**	1	0	0
	1	0	1	0
	2	0	0	1
	3	0	0	0

Cox Competing Risks – SAS

Summary of the Number of Event and Censored Values			
Total	**Event**	**Censored**	**Percent Censored**
438	27	411	93.84

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Without Covariates**	**With Covariates**
-2 LOG L	257.293	224.870
AIC	257.293	250.870
SBC	257.293	267.716

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	32.4225	13	0.0021
Score	33.2088	13	0.0016
Wald	29.4735	13	0.0056

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
manner	1	0.3192	0.5721
start	1	2.5463	0.1106
military	1	0.2422	0.6226
age	1	16.1192	<.0001
conflict	1	0.3055	0.5805
loginc	1	1.5105	0.2191
growth	1	1.0712	0.3007
pop	1	0.8676	0.3516
land	1	0.0497	0.8237
literacy	1	0.4876	0.4850
region	3	4.7021	0.1950

Cox Competing Risks – SAS

Analysis of Maximum Likelihood Estimates								
Parameter		**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**	**Label**
manner		1	0.37473	0.66325	0.3192	0.5721	1.455	
start		1	-0.05403	0.03386	2.5463	0.1106	0.947	
military		1	-0.36461	0.74091	0.2422	0.6226	0.694	
age		1	0.07386	0.01840	16.1192	<.0001	1.077	
conflict		1	-0.26085	0.47196	0.3055	0.5805	0.770	
loginc		1	0.32854	0.26732	1.5105	0.2191	1.389	
growth		1	0.08816	0.08518	1.0712	0.3007	1.092	
pop		1	0.00199	0.00214	0.8676	0.3516	1.002	
land		1	-0.0000397	0.0001781	0.0497	0.8237	1.000	
literacy		1	-0.00880	0.01260	0.4876	0.4850	0.991	
region	**0**	1	-0.65910	0.78519	0.7046	0.4012	0.517	region 0
region	**1**	1	-1.30176	0.80525	2.6134	0.1060	0.272	region 1
region	**2**	1	-1.43672	0.76049	3.5691	0.0589	0.238	region 2

Cox Competing Risks – R

```
cox_nat <- coxph(Surv(years, lost == "Natural") ~ manner + start
                 + military + age + conflict + loginc + growth
                 + pop + land + literacy + factor(region),
                 data = leaders)
summary(cox_nat)
```

Cox Competing Risks – R

```
## Call:
## coxph(formula = Surv(years, lost == "Natural") ~ manner + start +
##     military + age + conflict + loginc + growth + pop + land +
##     literacy + factor(region), data = leaders)
##
##   n= 438, number of events= 27
##    (34 observations deleted due to missingness)
##
##                       coef  exp(coef)   se(coef)      z Pr(>|z|)
## manner           3.747e-01  1.455e+00  6.633e-01  0.565    0.572
## start           -5.403e-02  9.474e-01  3.386e-02 -1.596    0.111
## military        -3.646e-01  6.945e-01  7.409e-01 -0.492    0.623
## age              7.386e-02  1.077e+00  1.840e-02  4.015 5.95e-05 ***
## conflict        -2.609e-01  7.704e-01  4.720e-01 -0.553    0.580
## loginc           3.285e-01  1.389e+00  2.673e-01  1.229    0.219
## growth           8.817e-02  1.092e+00  8.518e-02  1.035    0.301
## pop              1.991e-03  1.002e+00  2.138e-03  0.931    0.352
## land            -3.969e-05  1.000e+00  1.781e-04 -0.223    0.824
## literacy        -8.796e-03  9.912e-01  1.260e-02 -0.698    0.485
## factor(region)1 -6.427e-01  5.259e-01  8.360e-01 -0.769    0.442
## factor(region)2 -7.776e-01  4.595e-01  9.031e-01 -0.861    0.389
## factor(region)3  6.591e-01  1.933e+00  7.852e-01  0.839    0.401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cox Competing Risks – R

```
##                   exp(coef) exp(-coef) lower .95 upper .95
## manner               1.4546     0.6875   0.39644     5.337
## start                0.9474     1.0555   0.88657     1.012
## military             0.6945     1.4400   0.16255     2.967
## age                  1.0767     0.9288   1.03853     1.116
## conflict             0.7704     1.2980   0.30548     1.943
## loginc               1.3889     0.7200   0.82251     2.345
## growth               1.0922     0.9156   0.92423     1.291
## pop                  1.0020     0.9980   0.99780     1.006
## land                 1.0000     1.0000   0.99961     1.000
## literacy             0.9912     1.0088   0.96707     1.016
## factor(region)1      0.5259     1.9015   0.10217     2.707
## factor(region)2      0.4595     2.1763   0.07827     2.698
## factor(region)3      1.9330     0.5173   0.41484     9.007
##
## Concordance= 0.819  (se = 0.046 )
## Likelihood ratio test= 32.42  on 13 df,   p=0.002
## Wald test            = 29.47  on 13 df,   p=0.006
## Score (logrank) test = 33.21  on 13 df,   p=0.002
```

AFT Models with Competing Risks

- Accelerated Failure Time models have a similar structure to Cox regression models when dealing with competing risks.
- With AFT Models, distributions need to be evaluated for all types of failure!

CONDITIONAL PROCESSES

Independent Events?

- The competing risks approach presumes that each event type has its own hazard that governs **both** the occurrence and timing of events of that type.
- They are assumed to be independent processes acting in parallel with each other.
- Example:
  - Death due to natural causes vs. forcible removal from power.

Independent Events?

- In a business setting, this independence assumption rarely seems reasonable.
- Example:
  - Consider the event to be buying a personal computer.
  - Two types:
    - Mac
    - PC
  - These aren't two independent processes where we see what happens first.
- One process governs when you will buy a computer, while another process determines choice of computer.

Conditional Processes

- What if independence **DOES NOT** seem reasonable?
- **Conditional processes** occur when these events are NOT independent of each other – conditional on each other.
- Two Common Approaches:
  1. Two-Stage Modeling
  2. Fine-Gray Model

Conditional Processes

- What if independence **DOES NOT** seem reasonable?
- **Conditional processes** occur when these events are NOT independent of each other – conditional on each other.
- Two Common Approaches:
  1. Two-Stage Modeling
  2. Fine-Gray Model

Two-Stage Modeling

- **Two-Stage modeling** is when you model each of the stages (time and event) separately.
  1. **Survival analysis** – when will any event occur.
  2. **Classification model** – what type of event occurs.

- Example:
  1. Survival Analysis on buying a computer.
  2. Logistic regression on which type of computer to buy.