INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

22.04 – 10.05 2024 (3 ECTS)

Linus Andersson[1,2]

[1]University of Turku, [2]Stockholm University

linus.andersson@utu.fi

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

lecture V, Inference and extensions

Monday 06.05

Linus Andersson[1,2]

[1]*University of Turku*, [2]*Stockholm University*

linus.andersson@utu.fi

Ask not for whom the bell tolls…

- Ernest Hemingway

Inference and extensions

- We covered:
  - Non-parametric models
  - Parametric models
  - Data structure required for EHA

- Now
  - Assumptions and causality
  - Overview of extended EHA models
  - More on data structure required for EHA
  - Talk about the presentation, and the final report

Inference and extensions

Activity	Lesson topic	Keywords	Homework
Lesson 1	Introduction Key Concepts	Process time, Cencoring, Time-to-Event, Continous and discrete time	✓
Lesson 2	Key estimates Descriptive models	Kaplan-meyer, Density, Cum. Distribution function, Survival and Hazard funciton, Kaplan-Meyer, Life tables	✓
Lab 1. Non-Parametric models	Descriptive analysis		
Lesson 3	Key concepts, estimates for Parametric models	Exponential and Piece-wise exponential models, Shape parameter, the proportional hazard assumption, hazard ratios	✓
Lesson 4	Discrete and Continous models, Data structure	Time-varying variables, Cox,	
Lab 2. Parametric models I	Model fit		
Lesson 5	Piecing things together + Extensions	Competing risk models, Causality, heterogeneity	
Lab 3. Parametric models II	Diagnostics		
Lesson 6.	Discussion. Presentations.		Presentations

Inference and extensions

1. Assumptionsm, Inference and causality
2. Overview of extended EHA models
3. Data structure
4. Presentations and final report

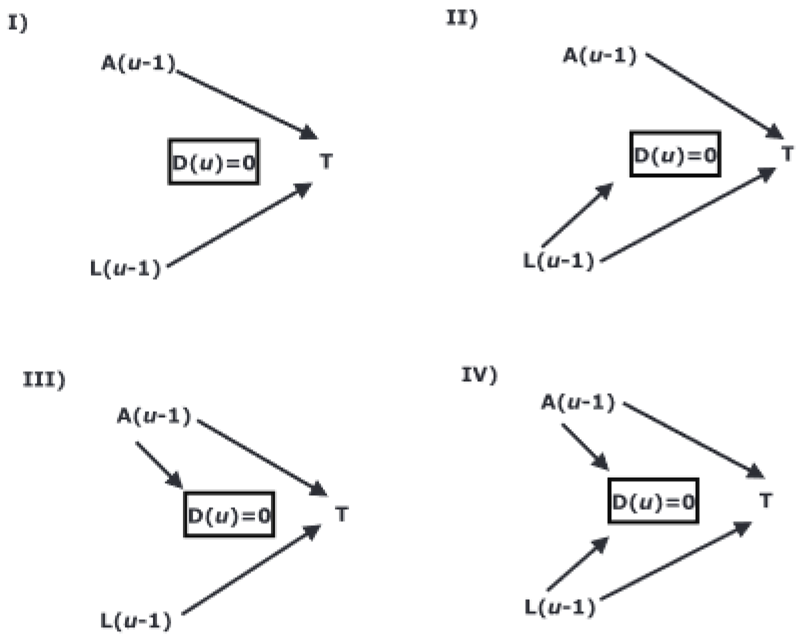
Assumptions, inference and causality

- A few scenarios to consider
    - Informative right censoring
    - Omitted variable bias
    - Duration effects on duration dependence
    - Mutually exclusive competing events - Competing risks

Informative censoring

- Informative censoring
  - Censoring correlated with event transition and/or predictors
  - Right censoring: ”attrition”, ”loss to follow-up”
  - Risk set after loss to follow-up no longer representative

- DAG of informative censoring (Ivanova 2013)
  - T = Duration to event transition
  - L(u) = Predictor
  - A(u) = Predictor
  - D(u) = loss to follow up

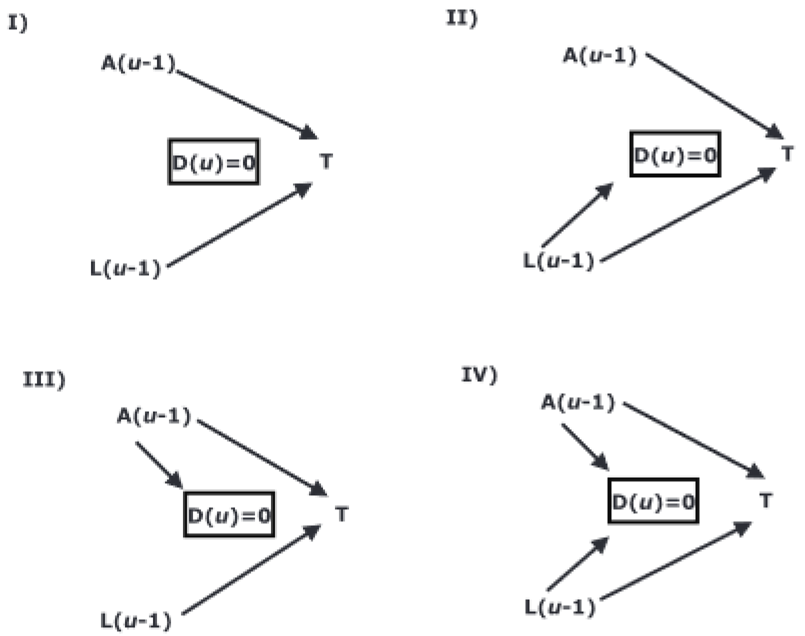


Howe et al (2016) Selection bias due to loss of follow up in cohor tstudies. Epidemiology, 2016, 27(1):91-97

Informative censoring

- Informative censoring considered in Ivanova et al (2013)
  - I) censoring random. Riskset before and after attrition has same properties. Not a source of bias
  - II) censoring correlate with having chidren, which predict lower h(t). Risk set changes. Bias.

- DAG of informative censoring
  - T = Duration to event transition (re-partnering)
  - L(u) = Have children
  - A(u) = Poor health
  - D(u) = loss to follow up

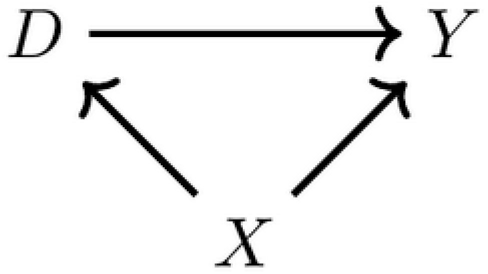


Howe et al (2016) Selection bias due to loss of follow up in cohor tstudies. Epidemiology, 2016, 27(1):91-97

Omitted variable bias

- Omitted variable bias
- Very broad term
- Unobserved heterogeneity
- Unobserved property results in a ”trajectory” of connected events across the life course

- DAG for Omitted variable bias
  - Y = Duration to event transition
  - D = predictor (observed)
  - X = propensity to event transition *and* predictor

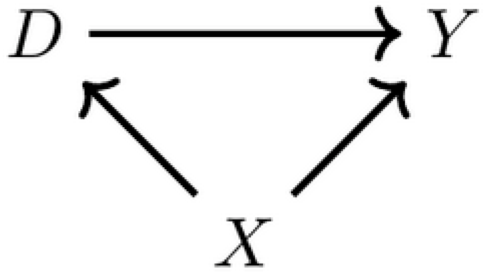


- To do

Omitted variable bias

- Omitted variable bias in Andersson (2019)
- Fertility intentions (X, unobserved) causes individuals to choose distance education (D) to accomodate fertility (Y).
- The effect of distance education on fertility is biased upward or entirely spurious

- DAG for Omitted variable bias
  - Y = Duration to event transition (first birth)
  - D = distance education (observed)
  - X = propensity to have a child (unbobserved



- To do

duration effects

- Duration effects on duration dependence
  - Duration dependence: the hazard s depend on the amount of time elapsed.
  - If hazard heterogeneous, high risk individuals (due to unobservables) X, will have events at earlier duration t than low risk individuals.
  - At higher duration t, risk set is depleted of high risk individuals
  - Population hazard will decrease at higher duration t

- Why is this a problem?
  - Positive duration depedence (higher risk at later t) understated
  - Negative duration dependence (lower risk at later t) overstated
  - Potentially falsely acredit hazard to duration dependence, rather than unobserved heterogeneity

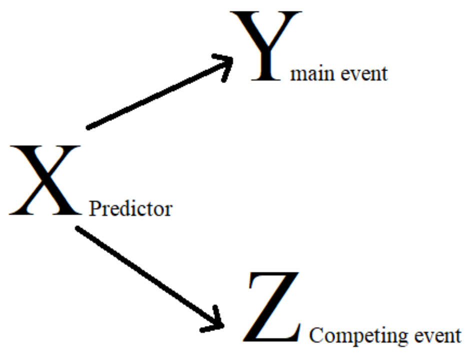
Proportional hazards assumption

- PH + Probability and timing = confusion
  - Groups may have different duration to event
  - But could also have the same probability of having a event
  - Stil, hazards will differ.
  - proportinal hazard assumption = hazard equal over time.
  - Unknown if hazard difference due to timing or probability
  - Solution: interaction with duration term (lab)

Competing events

- Competing events
  - Could be consided a case of informative censoring and censoring is known
  - Event that precludes the transition event of interest – a competing event!

- DAG for competing events
  - X predicts the main event Y
  - X also predicts a competing event, which precludes Y



- Problem
  - Coefficient of X on Y depend on the Coefficient of X on Z
  - The hazard of Y hard to interpret without somehow accounting for hazard of Z

Inference and extensions

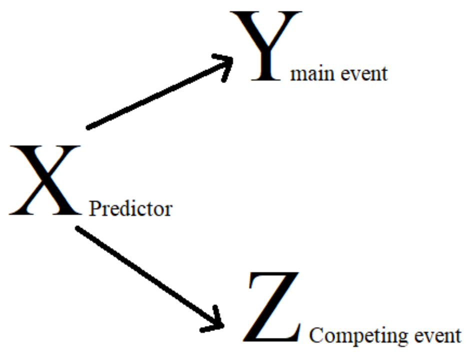
1. Assumptionsm, Inference and causality ✓
2. Overview of extended EHA models
3. Data structure
4. Presentations and final report

Overview of extended models

Model family	Used
Competing risk models	In the presence of competing risks
Frailty models	To account for unobserved heterogeneity across indivividuals in hazard, random effects.
Cure models	To account for unobserved heterogeneity (accounts for individuals in the risk set who are not ’actually’ at risk of event)

Competing risks models

- DAG for bias from competing events
  - Y = Main event
  - D = Competing event
  - X = Predictor



- Problem
  - Multiple mutually exclusive destinations states in state space
  - Individuals experience D cannot experience X
  - Risk of X is affected by risk of D

Howe et al (2016) Selection bias due to loss of follow up in cohor tstudies. Epidemiology, 2016, 27(1):91-97

Competing risks models

- Solution
  - Paramtric model with Separate cause specific likelihoods for competing events (continous time EHA)
  - Or, a Multinomial logistic regression (discrete time EHA)

- Estimate
  - Cause specific hazard
  - Cumulative Incidence function

- Key assumption
  - Censoring at competing event A is uncorrelated with risk of competing event B (if wasn't censored)
  - Often does not hold

Howe et al (2016) Selection bias due to loss of follow up in cohor tstudies. Epidemiology, 2016, 27(1):91-97

Unobserver heterogeneity – Competing risks

- DAG for bias from competing risks (Bernhardt et al 2016)
  - Y = Main event (leaving cohabitation to marriage)
  - D = Competing event (leaving cohabitation to single)
  - X = Predictor (Education)

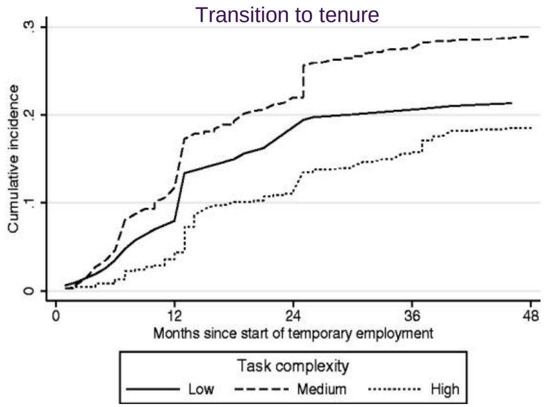
- Problem
  - Interested in one or more of multiple destination states
  - cohabitations who separate cannot marry
  - Risk of marriage is affected by risk of separation

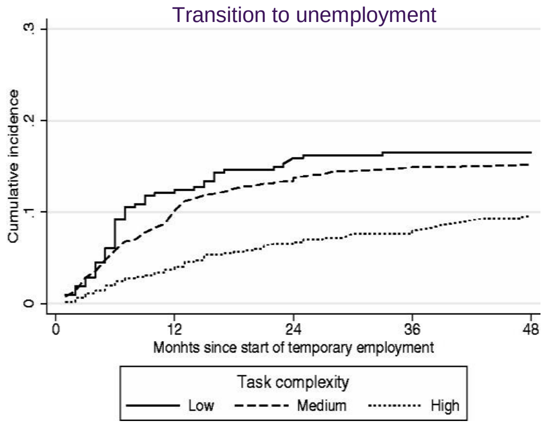
- Solution
  - Separate cause specific likelihoods for competing events
  - “Instantaneous risk of experiencing event j at *t* under the condition neither j and nor any other event has occurred till *t*”

Howe et al (2016) Selection bias due to loss of follow up in cohor tstudies. Epidemiology, 2016, 27(1):91-97

Competing risks models

- Example: transition from temporary to tendured employment or unemployment
  - Tenured employment and unemployment are mutually exclusive censoring events
  - CIF – cumulative failure incidence





Eur Sociol Rev, Volume 31, Issue 5, October 2015, Pages 558–572, https://doi.org/10.1093/esr/jcv055

Inference and extensions

1. Assumptionsm, Inference and causality ✓
2. Overview of extended EHA models ✓
3. Data structure
4. Presentations and final report

15 minute break

Data structure

```
stset studytime, failure(died) id(id)
```

	id	died	studytime	transplant	wait_to_transplant
1	2	1	6	0	0
2	38	1	5	1	3

- Above, a convenient structure. Not allways the case
  - How was entry at risk, studytime = 0, derived?
  - How was the window of observation accounted for?

Data structure

id	event	loss	censor_event	censor_loss	win_start	win_end	age_10	age_20	end
1	1	0	1975m3	.	1980m1	2000m1	1965m9	1975m9	1975m3
2	1	0	1990m5	.	1980m1	2000m1	1985m2	1995m2	1990m5
3	0	1	.	1983m3	1980m1	2000m1	1975m11	1985m11	1983m3
4	1	0	2010m10	.	1980m1	2000m1	1995m4	2005m4	2000m1
5	1	0	2005m2	.	1980m1	2000m1	2003m1	2013m1	2000m1
6	1	0	1982m1	.	1980m1	2000m1	1985m7	1995m7	1982m1

- For example. variables for the time id
  - Censored at event, loss of follow-up (here censor_event, censor_loss)
  - Comes under observation, Is no longer under observation (here: **`win_start, win_end`**)
  - Enter risk, exist risk (here: **`age_10, age_20`**)
  - The earliest end occurence of above events – (here: **`end = min(censor_loss censor_event win_end)`**)

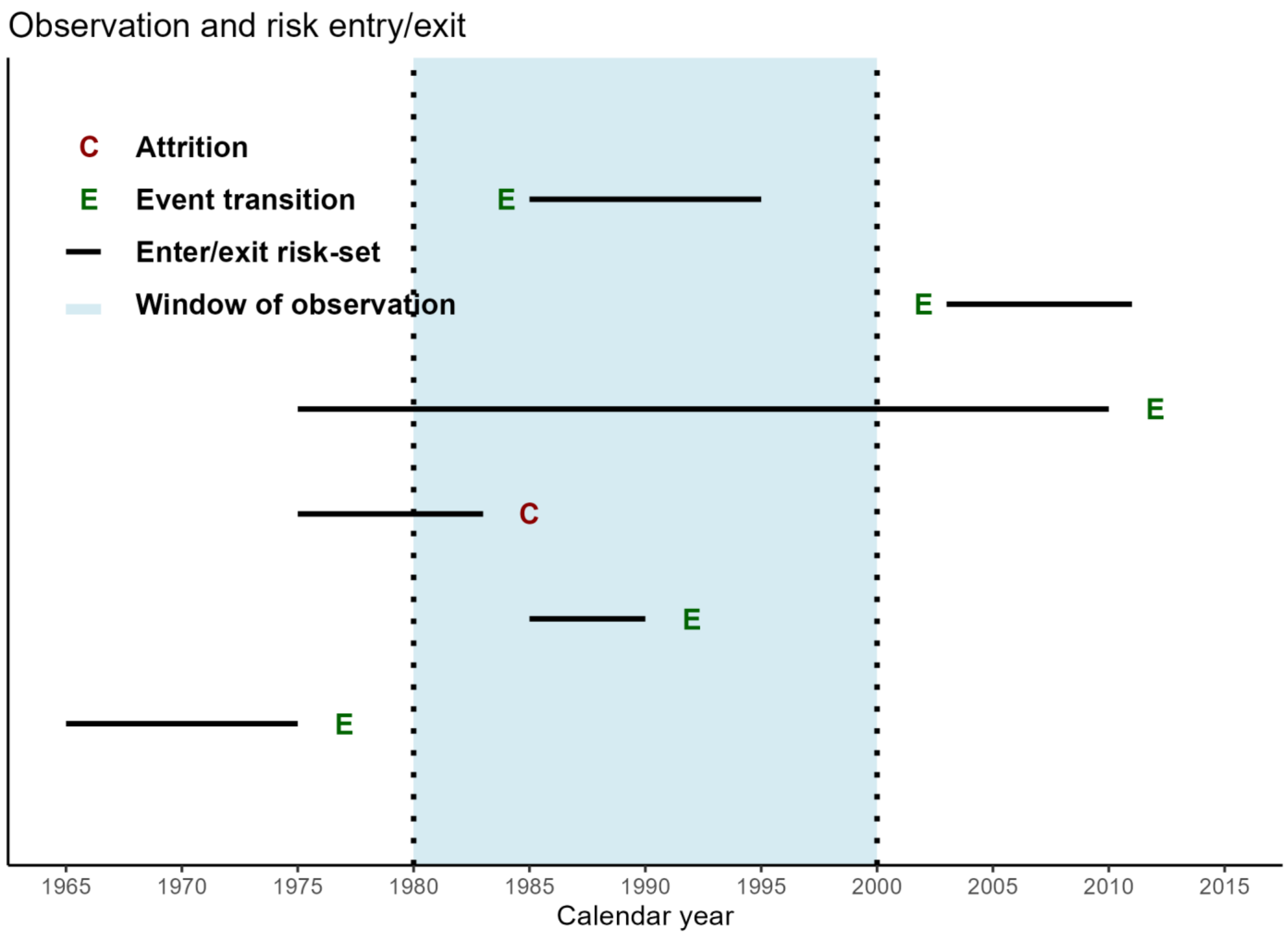
Data structure

```
stset end, fail(event) origin(time age_10) enter(time win_start)
```

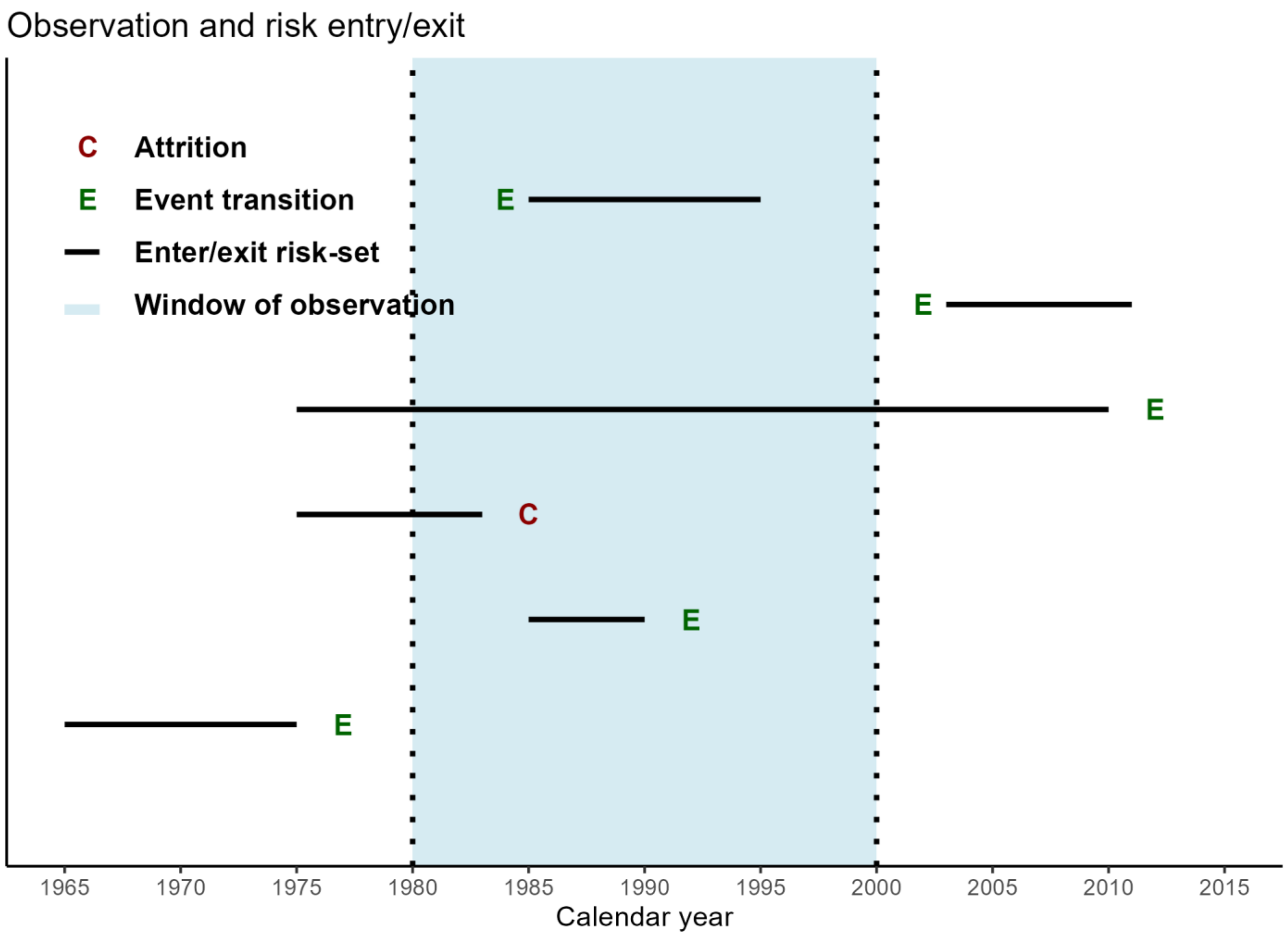
id	event	loss	censor_event	censor_loss	win_start	win_end	age_10	age_20	end
1	1	0	1975m3	.	1980m1	2000m1	1965m9	1975m9	1975m3
2	1	0	1990m5	.	1980m1	2000m1	1985m2	1995m2	1990m5
3	0	1	.	1983m3	1980m1	2000m1	1975m11	1985m11	1983m3
4	1	0	2010m10	.	1980m1	2000m1	1995m4	2005m4	2000m1
5	1	0	2005m2	.	1980m1	2000m1	2003m1	2013m1	2000m1
6	1	0	1982m1	.	1980m1	2000m1	1985m7	1995m7	1982m1

- For example. variables for the time id
  - Censored at event, loss of follow-up (here censor_event, censor_loss)
  - Comes under observation, Is no longer under observation (here: `win_start, win_end`)
  - Enter risk, exist risk (here: `age_10, age_20`)
  - The earliest end occurence of above events – (here: `end = min(censor_loss censor_event win_end)`)

Data structure

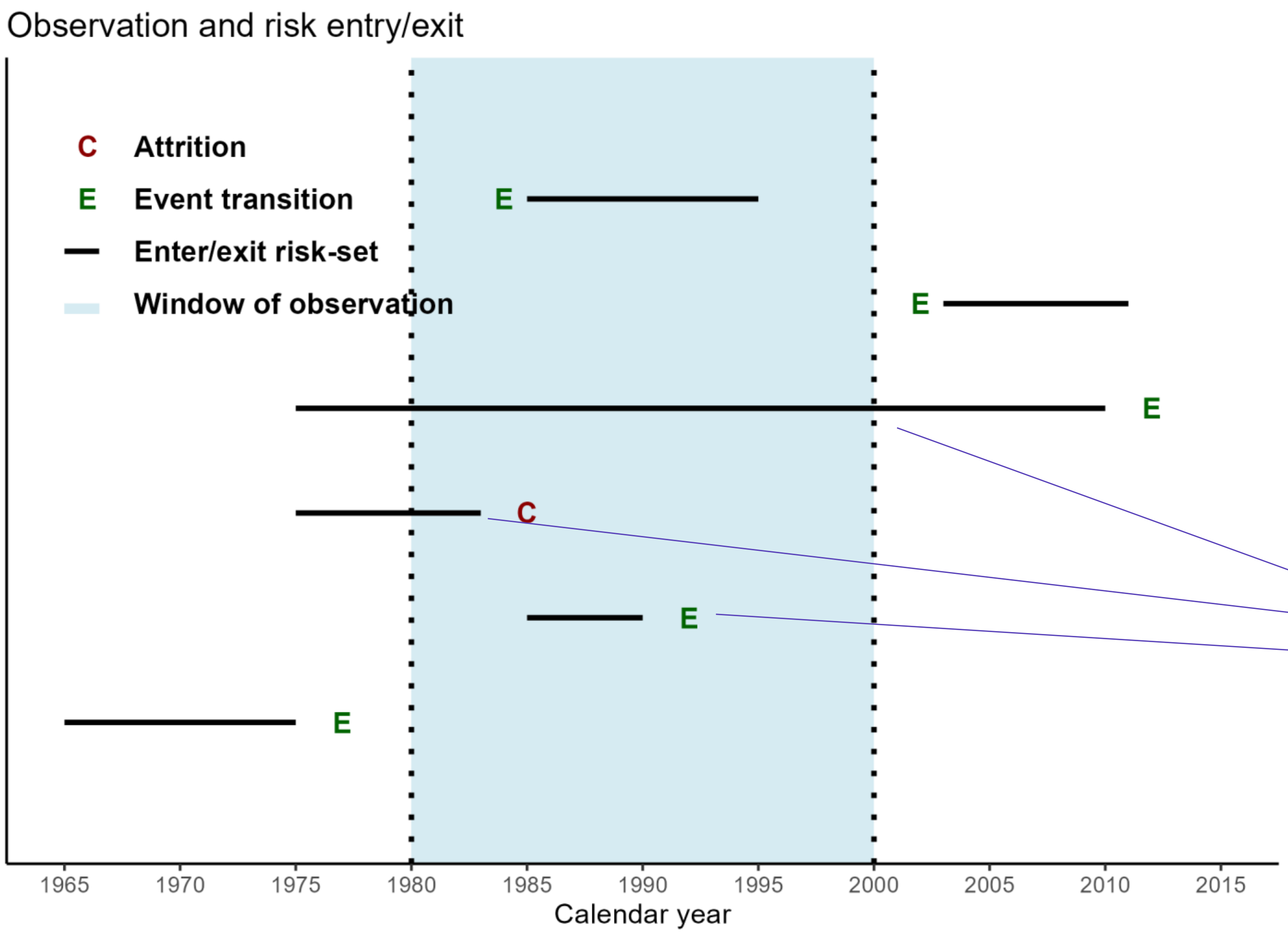


Data structure



id	event	censor_event	win_start	win_end	age_10	age_20
6	1	1982m1	1980m1	2000m1	1985m7	1995m7
5	1	2005m2	1980m1	2000m1	2003m1	2013m1
4	1	2010m10	1980m1	2000m1	1995m4	2005m4
3	0	.	1980m1	2000m1	1975m11	1985m11
2	1	1990m5	1980m1	2000m1	1985m2	1995m2
1	1	1975m3	1980m1	2000m1	1965m9	1975m9

RECAP LESSON i and II - Continous and discrete time



id	event	censor_event	win_start	win_end	age_10	age_20
6	1	1982m1	1980m1	2000m1	1985m7	1995m7
5	1	2005m2	1980m1	2000m1	2003m1	2013m1
4	1	2010m10	1980m1	2000m1	1995m4	2005m4
3	0	.	1980m1	2000m1	1975m11	1985m11
2	1	1990m5	1980m1	2000m1	1985m2	1995m2
1	1	1975m3	1980m1	2000m1	1965m9	1975m9

id	_st	_d	_origin	_t	_t0
6	0	.	306	.	.
5	0	.	516	.	.
4	1	1	423	57	0
3	1	0	190	88	50
2	1	1	301	63	0
1	0	.	68	.	.

Inference and extensions

1. Assumptionsm, Inference and causality ✓
2. Overview of extended EHA models ✓
3. Data structure ✓
4. Presentations and final report

Presentation and final report

- Presentation
  - 5 min
  - Introduce your research project for the report
  - Motivate the use of EHA
  - Describe data
  - Describe ideal data
  - Interpret results (if any)

Presentation and final report

- Report
  - Describe the research question
  - Motivate the use of EHA
  - Describe data
  - Fit EHA models
  - Interpret results
  - Discuss ideal research design
  - Further Instructions in Moodle

Inference and extensions

1. Assumptionsm, Inference and causality ✓
2. Overview of extended EHA models ✓
3. Data structure ✓
4. Presentations and final report ✓