INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

22.04 – 10.05 2024 (3 ECTS)

Linus Andersson[12]

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

linus.andersson@utu.fi

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

Lecture I – Course introduction & Key concepts of EHA

Monday 22.04

Linus Andersson[12]

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

linus.andersson@utu.fi

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

Lecture I, part I – Course introduction

Monday 22.04

Linus Andersson[12]

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

linus.andersson@utu.fi

> Failure to prepare is preparing to fail
>
> - Scott Tennant

Course overview

1. Introduction
2. Learning outcomes
3. Workflow
4. Examination

Hello!

- Quick introduction (1-2 minute)
  - Name
  - Study dicipline
  - Research or study interests if any

Course overview

1. Introduction ✓
2. Learning outcomes
3. Workflow
4. Examination

Learning outcomes

- Understand the logic of EHA
- Design a basic event history study to examine a research question
- Apply basic event history analysis in Stata
- Interpret results and draw conclusions from EHA

Course overview

1. Introduction ✓
2. Learning outcomes ✓
3. Workflow
4. Examination

workflow

Report/ HW	Topic	Lesson focus
Good choices From RQ to Data to Model	Key concepts of EHA: Censoring; Continous and discrete time; The What and Why of EHA; "Building blocks of time"	Fluency, intuition
	Key estimates in EHA: Hazard function; Survival function	By-hand calculation
	Descriptive Key estimates in EHA - Non parametric models: Life tables; Kaplan-Meier	Data prep, Stata
	Model EHA - Semi-parametric and Parametric models: Model choice; Estimation; Diagnostics; Interpretation	Data prep, Intuition, Stata

Lectures, labs and activities

- 6 lectures inc. stata walkthroughs
- 3 Labs with focus on Stata
- 3 Homework assigments
- Presentation of report plan at final lecture
- Report

Scheduling

Date	Time	Activity	Lesson topic	Keywords	Homework
Mon 22.04	12:15-14:00	Lesson 1	Introduction Key Concepts	Process time, Cencoring, Time-to-Event, Continous and discrete time	✓
Wed 24.04	10:15-12:00	Lesson 2	Key estimates Descriptive models	Kaplan-meyer, Density, Cum. Distribution function, Survival and Hazard funciton, Kaplan-Meyer, Life tables	✓
Fri 26.04	10:15-12:00	Lab 1. Non-Parametric models	Descriptive analysis		
Mon 29.04	12:15-14:00	Lesson 3	Key concepts, estimates for Parametric models	Exponential and Piece-wise exponential models, Shape parameter, the proportional hazard assumption, hazard ratios	✓
Thu 02.05	10:15-12:00	Lesson 4	Discrete and Continous models, Data structure	Time-varying variables, Cox, Logit	
Fri 03.05	10:15-12:00	Lab 2. Parametric models I	Model fit		
Mon 06.05	12:15-14:00	Lesson 5	Piecing things together + Extensions	Case studies with focus on model choice, Interpretation ation of causality, heterogeneity	
Wed 08.05	10:15-12:00	Lab 3. Parametric models II	Diagnostics	Process time, Cencoring, Time-to-Event, Continous and discrete time	
Fri 10.05	10:15-12:00	Lesson 6.	Discussion. Presentations.	Kaplan-meyer, Density, Cum. Distribution function, Survival and Hazard funciton, Kaplan-Meyer, Life tables	Presentations

Online etiquette

- Unmute and interrupt me if you have a question.
- Please do no use chat box or emoji's for this – I will not notice in time
- Video on/off is you choice – I will assume that you are there.
- We will have 15 minute breaks to recharge
- During breaks, I will not be in front of keyboard. I'll be preparing for the lectures or have a break.
- During breaks, unmute if you want to stay and socialize.

Course overview

1. Introduction ✓
2. Learning outcomes ✓
3. Workflow ✓
4. Examination

Examinations and required hand-ins

- Requried hand-ins:
  - All homework
  - Report
- Required participation
  - 3 out of 6 lectures
- The labs are not mandatory
  - But highly recommended

Report

- Report
  - Real or hypothetical analysis
  - 1-2 pages, 1 line spacing
  - RQ & motivation
  - Describe data, methods
  - Analyze and interpret
  - Remember: this is an exercise, not a large project
- Data for report
  - Own data, or
  - training data

Course overview - summary

- At this point, make sure to
  - Orient yourself in Moodle
  - Get you mind set on doing some homework
  - Start thinking about report (we will return to this throughout the course)

Course overview

1. Introduction ✓
2. Learning outcomes ✓
3. Workflow ✓
4. Examination ✓

15 minute break

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

fluxconsortium.fi

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

Lecture I, part II – Key concepts of HA

Monday 22.04

Linus Andersson[12]

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

linus.andersson@utu.fi

fluxconsortium.fi

Course overview

1. The What and Why of event history analysis
2. The building blocks of time
3. Censoring
4. Continous vs discrete time

Some literature for inspiration

	Event transition	Broader question / time process
Skopek et al 2020	PhD completion	How does Universities (EUI) influence PhD completion?
Sherkat 1991	Religious faith conversion	What factors predicts religious conversion over the life course?
Robinsson & Smit-Lovin 2001	Humor in Conversations	What is the structure of conversations? what predicts the use of humor in conversation
Ivanova et al 2013	Finding a partner after divorce	Does having a child prevent women from finding a partner after union separation?
Drefahl et al 2020	COVID mortality	What factors influence COVID mortality in Sweden?
De Mesquita & Smith 2010	Political Leaders duration in power	How do elected/non-elected leaders stay in power?
Blossfeld & Huinink 1991	Becoming a parent	Why do highly educated women have children later than less educated women?
Asadulla et al 2021	Having a son vs having a daughter	Are there still Son preferences an Bangladesh?
Andersson 2018	Fertility	Is there an effect of studying online for childbearing behavior?
Alsan & Goldin 2019	Child mortality	On the relationship between child mortality and installing clean water and sewage in 1800-1900 US

See Moodle for a complete literature reference list and links to all articles

Names differ across diciplines

- Many names
  - Event history analysis (sociology & demography)
  - Survival analysis (epidemiology)
  - Duration/time-to-event analysis (economics)
  - Reliability analysis (engineering)

- The units of analysis and outcomes also vary across dicipline
  - Therefor multiple terms for the same thing circulate
  - Often non-intuitive terms (”failure”, ”censoring”, ”the risk of marriage”)
  - Do not ponder to much on the wording – accept, learn and use as lingo

EHA - a analysis of how history explain events

- Event
  - Something happen. A transition from state A to B. Alive to Dead.
- History
  - The transition result from a process unfolding over time.
- Analysis
  - Descriptive and causal analysis
  - but not a technique for causal identification

Time - the unit of analysis

- Linear regression – metric data
- Logistic regression – binary data
- Poisson regression – count data
- EHA regression – Time-to-event data

Time - the unit of analysis

Example using a fotball game

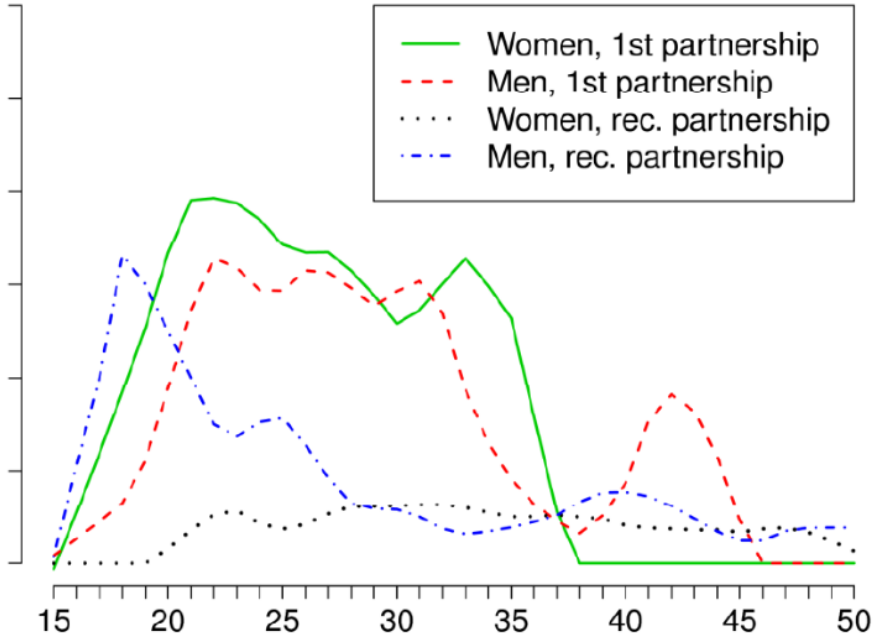
- Linear regression – metric data
  - Y = Over-time played.
- Logistic regression – binary data
  - Y = home-team victory.
- Poisson regression – count data
  - Y = nr of goals scored.
- EHA regression – Time-to-event data
  - Y = The duration to the event of first goal.

When to use event history analysis

- EHA appropriate when
  1. When Y is qualitative event, not quantitative
  2. We are interested in an time-dependent process leading up to Y
  3. We care about when Y occur as well as whether Y occur

What is a time-dependent process?

- Example: the time process of transitions into 1st & 2nd partnership
  - The events below do not randomly occur over time
  - The events occur according to pattern across time
  - This pattern can be of interest in itself
  - The predictor of events may occur over time
  - The influence of the predictor on events may be time-dependent.



What is a time-dependent process?

- Example: the time process of transitions into 1st & 2nd partnership
  1. Partnership formation increases, then decreases, over age
  2. Employment status vary over the ages
  3. Employment status influences partnership formation
  4. And differentially so at different ages

When and whether as reseach questions

- University leadership on phd completion (Skopek, Triventi, Blossfeld 2020)
  - Here, we care about what causes someone complete their phd or not
  - But also how long students take to complete their phd

Recap - when to use event history analysis

- EHA appropriate when
  1. When Y is qualitative, not quantitative
  2. We are interested in an time-dependent process leading up to Y
  3. We care about when Y occur as well as whether Y occur

Try in vain to answer these questions without EHA!

- A ”when” question using linear regression
  - RQ: How does women’s educational attainment influence age of motherhood?
  - Y = time to first birth (continuous)
  - X = educational level

- Some problems here
  - So, do highly educated women prefer childbearing at later ages, or childbearing as soon as they are done with their education, regardless of level?
  - You have a time-dependent process and non-occurence to deal with AND issues with substantive interpretation

Try in vain to answer these questions without EHA!

- But check yourself:
  - Could be modelled with EHA != Should be modelled with EHA
  - Do not use EHA for every conceivable question!

Recap - The what and why of event history analysis

- Summary
  - Is the RQ When and Whether?
  - Are we dealing with a time-dependent processes?
  - Do we need to deal with censoring?

- If this setup define the questions you want to examine, then EHA might be something for you!

Course overview

1. The What and Why of event history analysis ✓
2. The building blocks of time
3. Censoring
4. Continous vs discrete time

The building blocks of time to event

- State space (what are the possible states)
  -
- Transition event (from origin state and destination state)
  -
- Process time (define what time to measure)
  -
- Episode/spell (duration in a particular state)
  -
- At risk condition (who can experience the event? – the risk set)
  -
- Process time ends
  -

The building blocks of time to event of death

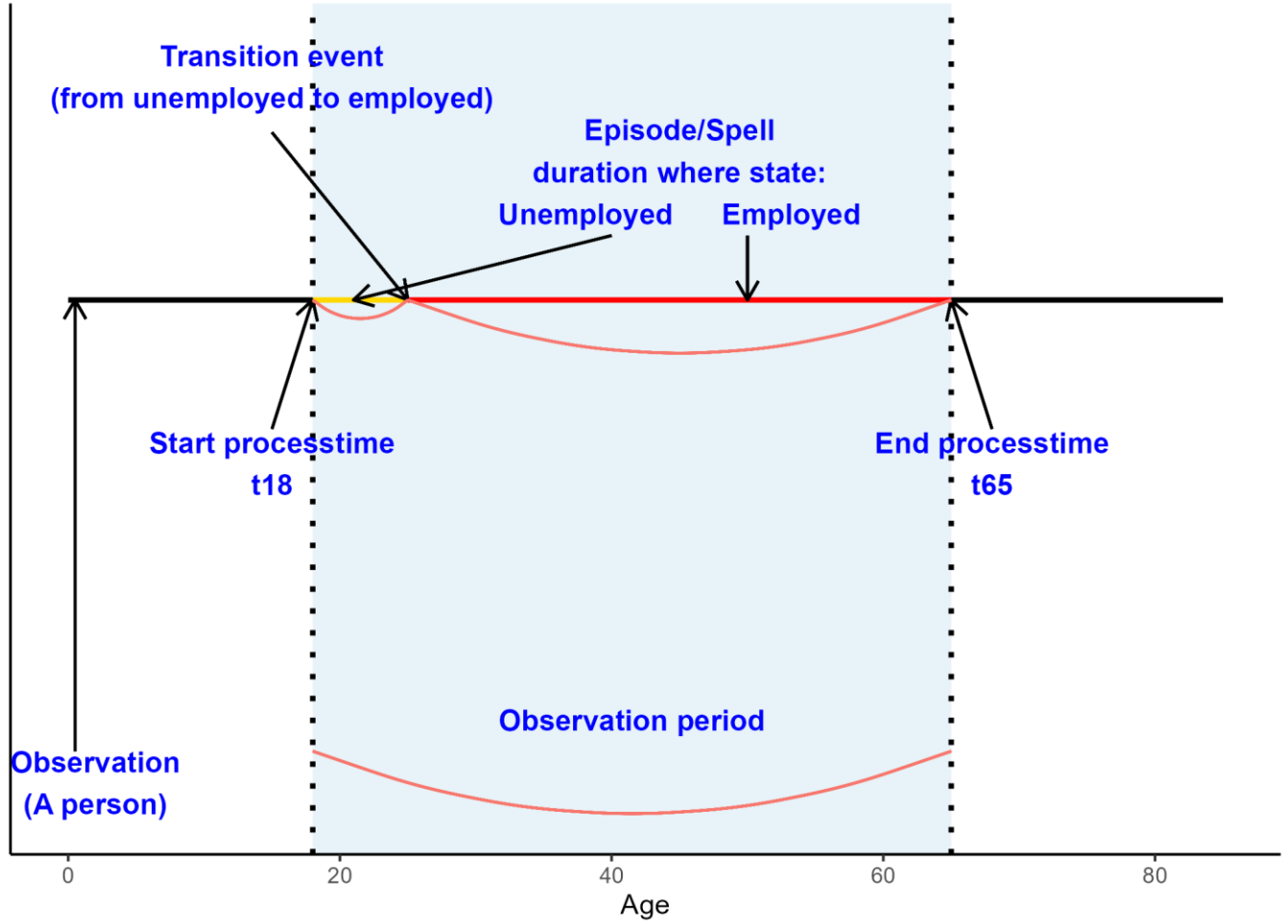
- State space (what are the possible states)
  - Alive, dead
- Transition event (from origin state and destination state)
  - From alive to dead
- Process time (define what time to measure)
  - Since age 0?
- Episode/spell (duration in a particular state)
  - Time spent alive (age)
- At risk condition (who can experience the event? – the risk set)
  - Alive
- Process time ends
  - death, cencoring.

The building blocks of time to event of first employment

- State space (what are the possible states)
  - Empoyed, other
- Transition event (from origin state and destination state)
  - From non-employed to employed
- Process time (define what time to measure)
  - Since age 18?
- Episode/spell (duration in a particular state)
  - The time spent in outside employment
- At risk condition (who can experience the event? – the risk set)
  - Not in employment
- Process time ends
  - first employment, cencoring.

The building blocks of time to event of death

- Here a g



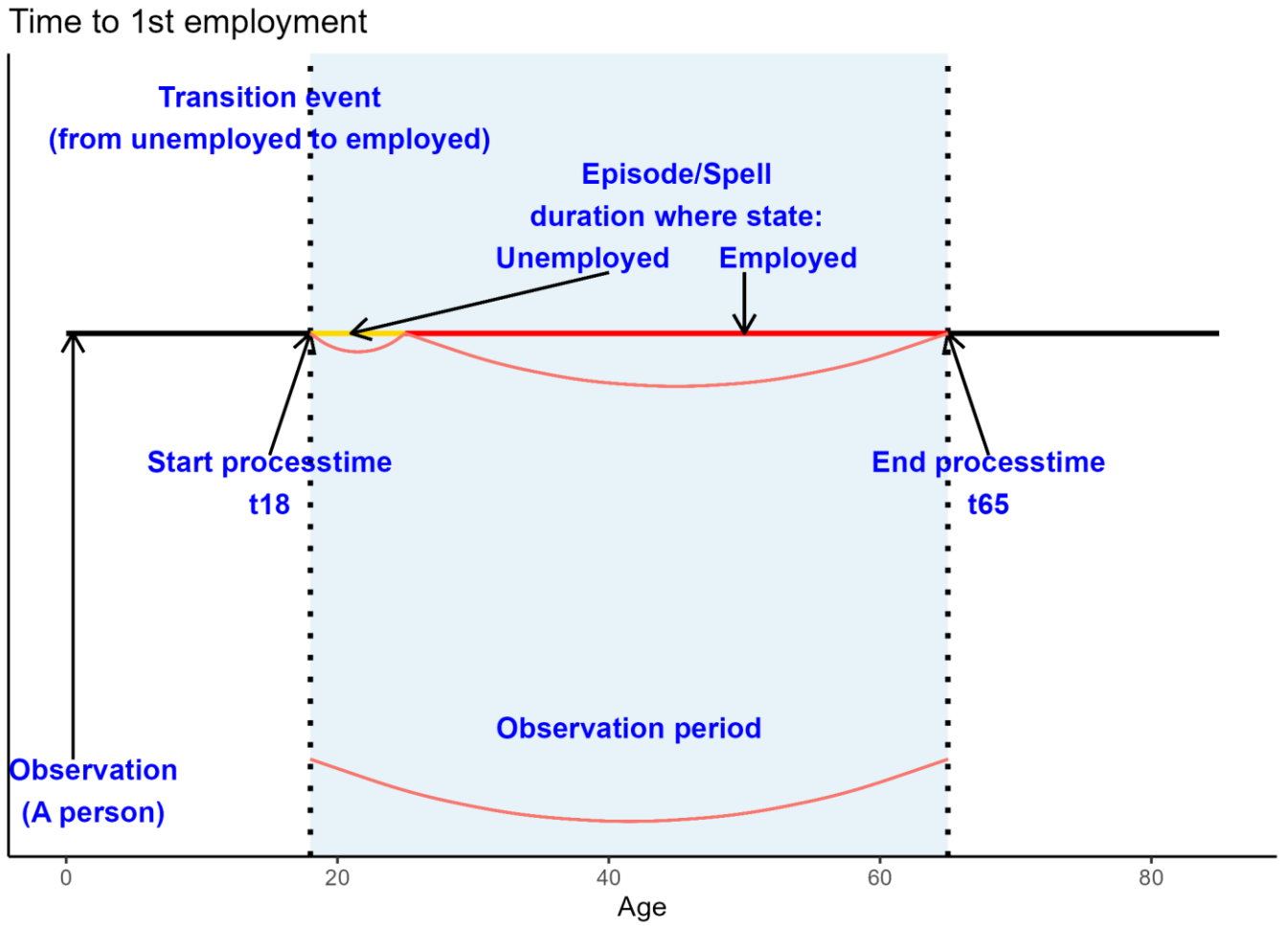
Key concepts of EHA

1. The What and Why of event history analysis ✓
2. The building blocks of time ✓
3. Censoring
4. Continous vs discrete time

Censoring

- EHA has a particular class of issue with missing data points
  - Incompete data at time t
- Incompete data at time t changes the risk set → changes our estimates
  - We will se how later
- The impact on estimates depends on
  - When in the time-process we have incomplete data
  - The cause for incomplete data at time t
- Cencoring & Truncation - Dealing with incomplete data at time t
  - ”cencored event times”, ”censor at event”

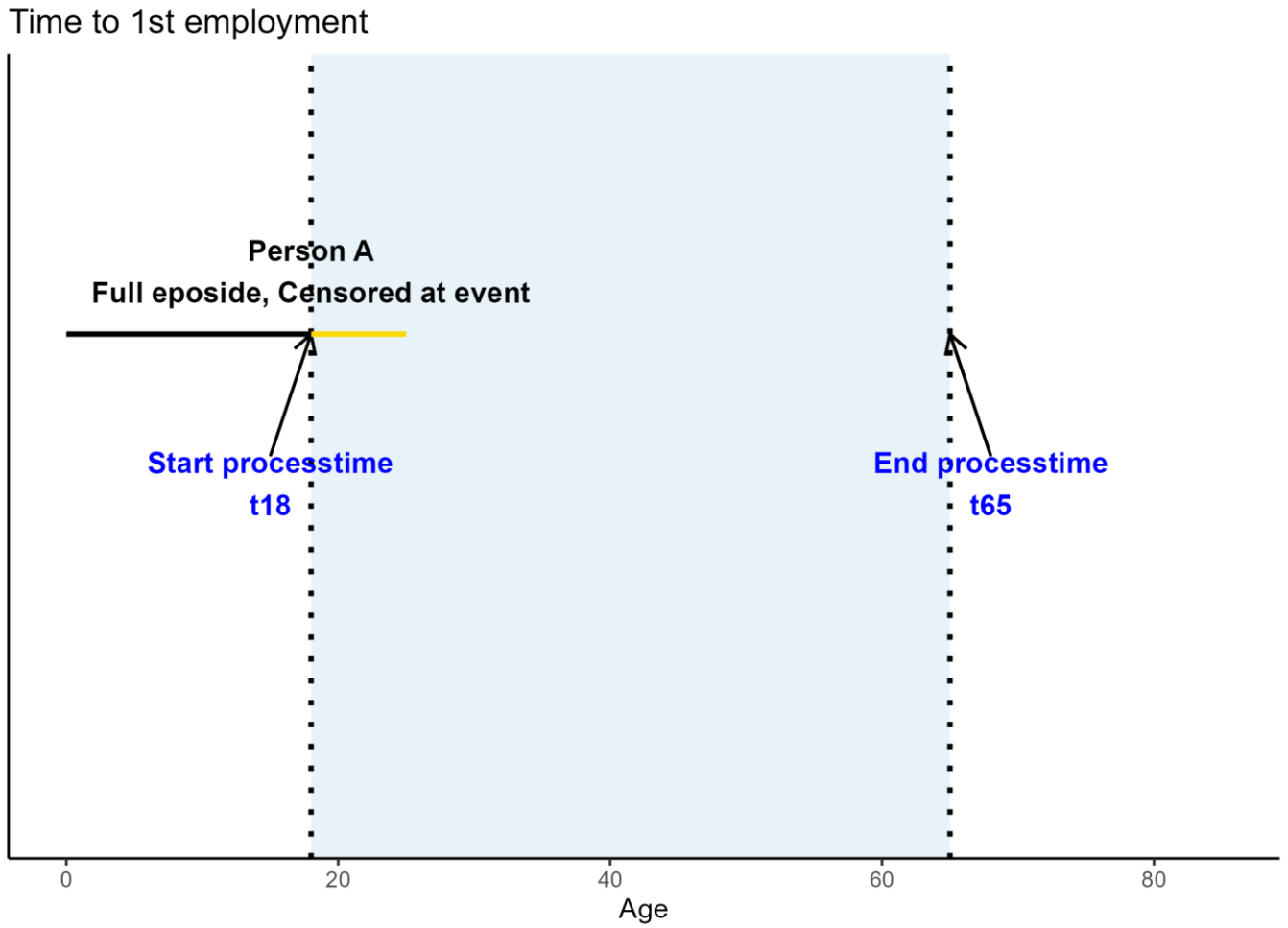
Censoring


Time to 1st employment
Transition event
(from unemployed to employed)
Episode/Spell
duration where state:
Unemployed
Employed
Start processtime
t18
End processtime
t65
Observation period
Observation
(A person)
0
20
40
60
80
Age

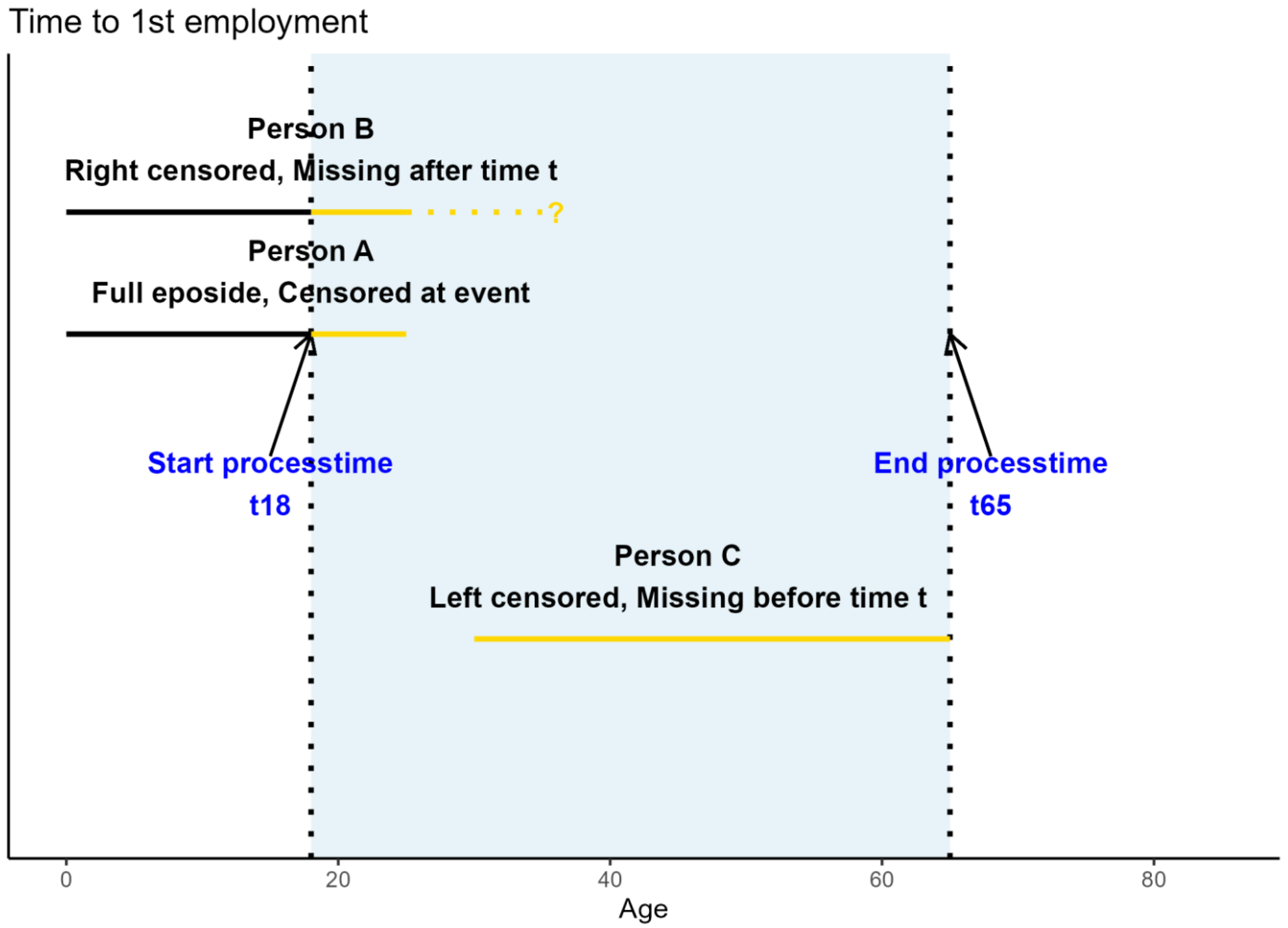
Censoring


Time to 1st employment
Start processtime
t18
End processtime
t65
(A person)
0
20
40
60
80
Age

Censoring


Time to 1st employment
Person A
Full eposide, Censored at event
Start processtime
t18
End processtime
t65
0
20
40
60
80
Age

Censoring


Time to 1st employment
Person B
Right censored, Missing after time t
?
Person A
Full eposide, Censored at event
Start processtime
t18
End processtime
t65
Person C
Left censored, Missing before time t
0
20
40
60
80
Age

Censoring

- Not problematic censoring (un-biased estimates)
  - Right censoring: reason for censoring (e.g. attrition) is uncorrelated with the risk of having had the event
  - Left censoring: reason for censoring is uncorrelated with the risk of having the event in the future net of controls (right censoring) (”Non-uninformative”)
- Very problematic censoring (biased estimates)
  - Above assumptions violated: If reason for cencoring is correlated with the event (”informative”)
  - Use critical thinking, previous knowledge.
- Intuition
  - We apply cencoring to adjust the risk set – the denominator.
  - So if cencoring is (not) random then the risk set is (not) a random sample after adjustment

Censoring

- Subtle differences
  - Truncated data = Exposure time not observed
  - Censored data = The Event transition is not observed

Recap - censoring

- Different kinds of censoring
- Dealing with censoring is a Key advantage of EHA
- You need to define censoring given the RQ, process-time and the data at hand.

Key concepts of EHA

1. The What and Why of event history analysis ✓
2. The building blocks of time ✓
3. Censoring ✓
4. Continous vs discrete time

Continous and discrete time EHA

Doing EHA involves at least three choices

1. Concieve of a process as Continous or discrete time
2. consider hat is the smallest data time-unit (interval) at hand
3. Use Contionus or discrete EHA models

- Step 1 & 2, defining data and process as continous or discrete:
  - Ultimately has a arbitrary or theoreticall elements.
- But step 3, continous and discrete EHA models:
  - Have actual non-arbitrary differences in mathematical properties.

Continous and discrete time – 2 examples

Continous time: The event of conception

- Can occur any time - From year, month, to infinity – a continous process
- We have monthly level data, and consider it adequate.
- We fit a continous EHA model, geting a montly hazard function (more on this later)

Continous and discrete time – 2 examples

”As if discrete time”: The event of conception with less precise data

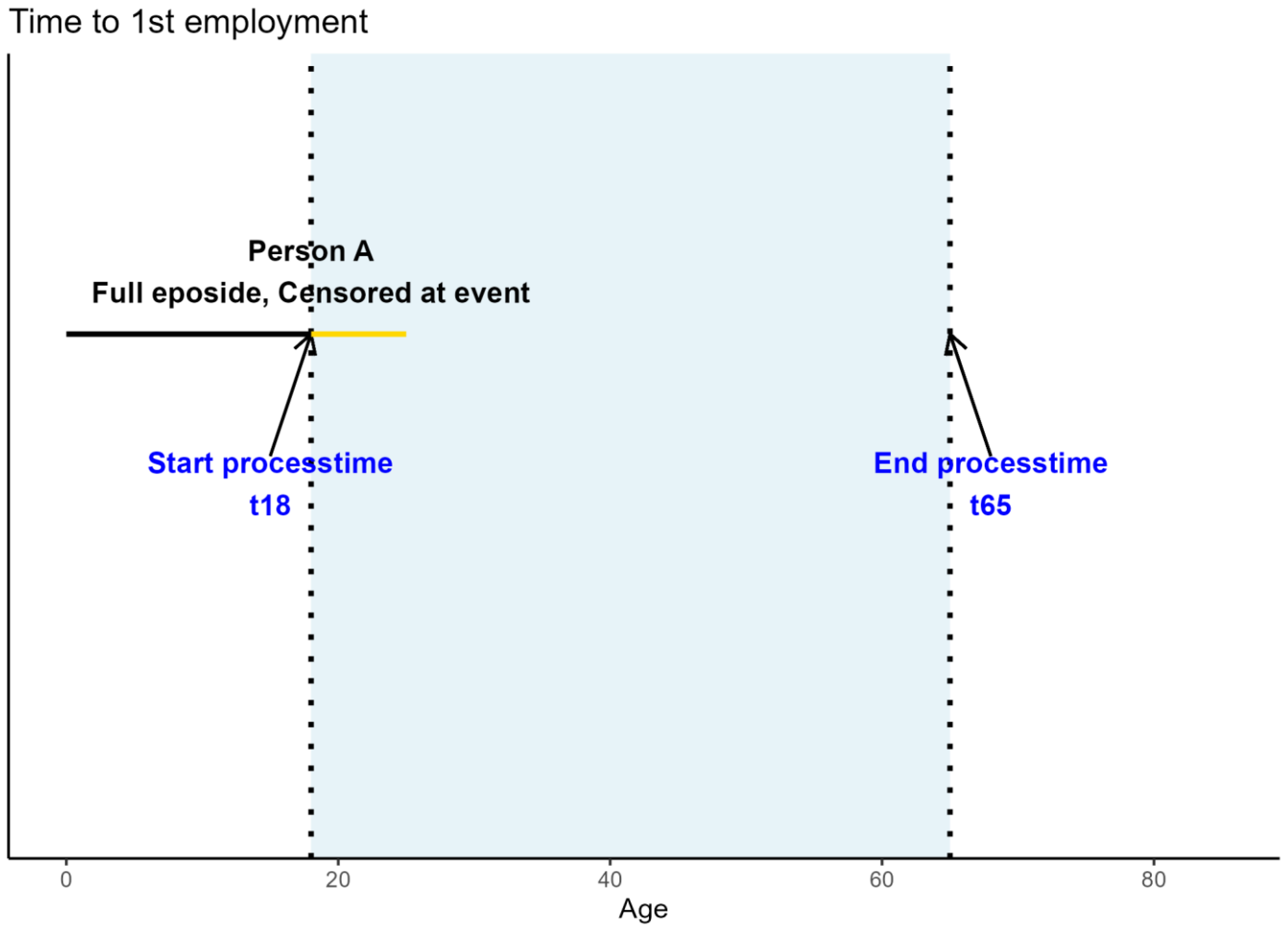
- Can occur any time - From year, month, to infinity – a continous process
- But we just have data on a yearly basis
- We fit a discrete EHA model, estimating a probability of conception during a yearly intervall spell, conditional on no event prior to this interval, assuming the risk of event is equal in any time during this interval spell (more in this later)

Continous and discrete time – 2 examples

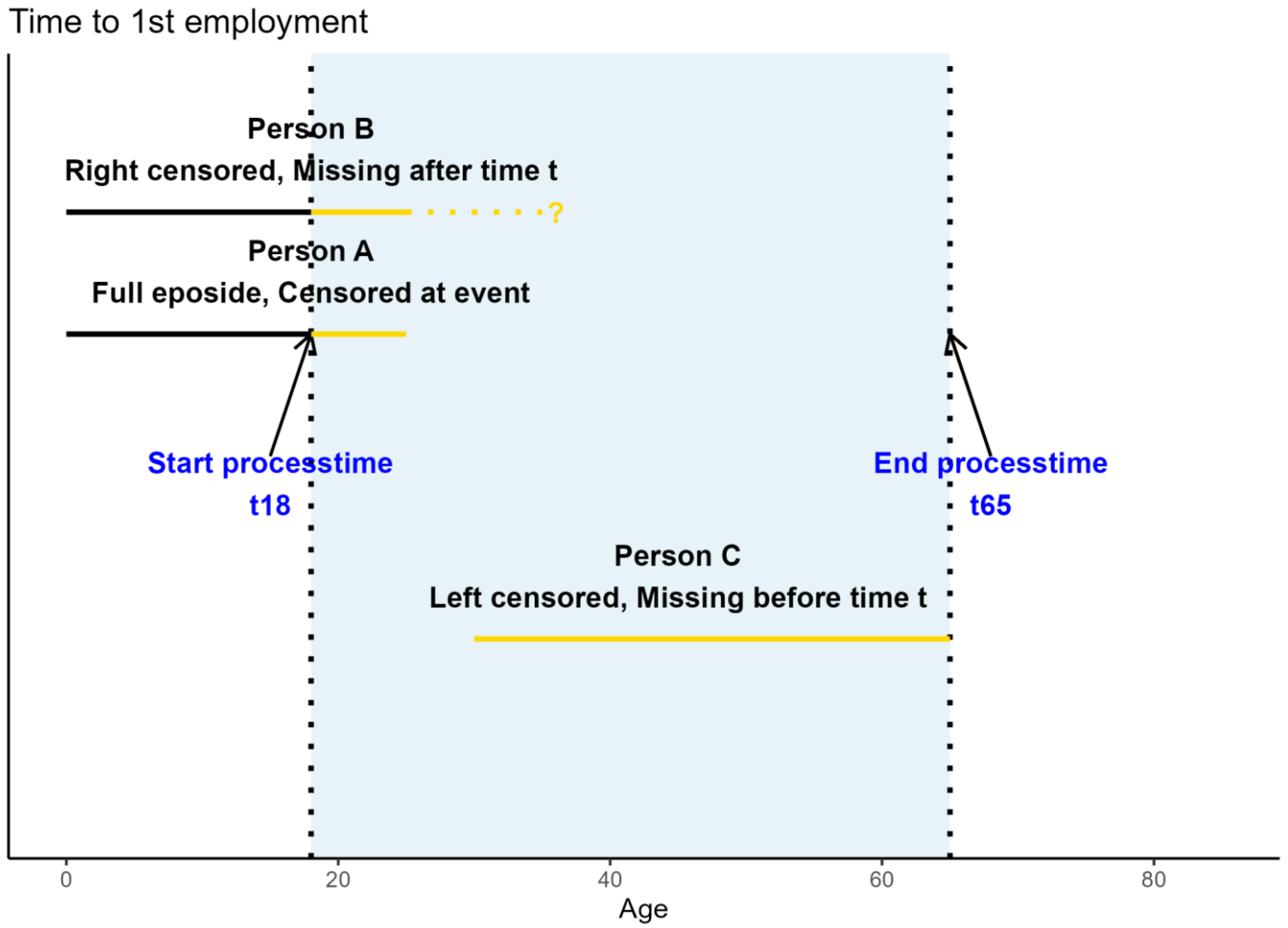
”Truly discrete time”: The event of entry into medical school

- Can occur in any given year - but not months (only in september) – a discrete process
- We reason that data in years is adequate
- We have to fit a discrete EHA model, estimating a probability of enrolling in medical school during a yearly intervall spell t, conditional on not having yet enrolled until this point intervall (more in this later)

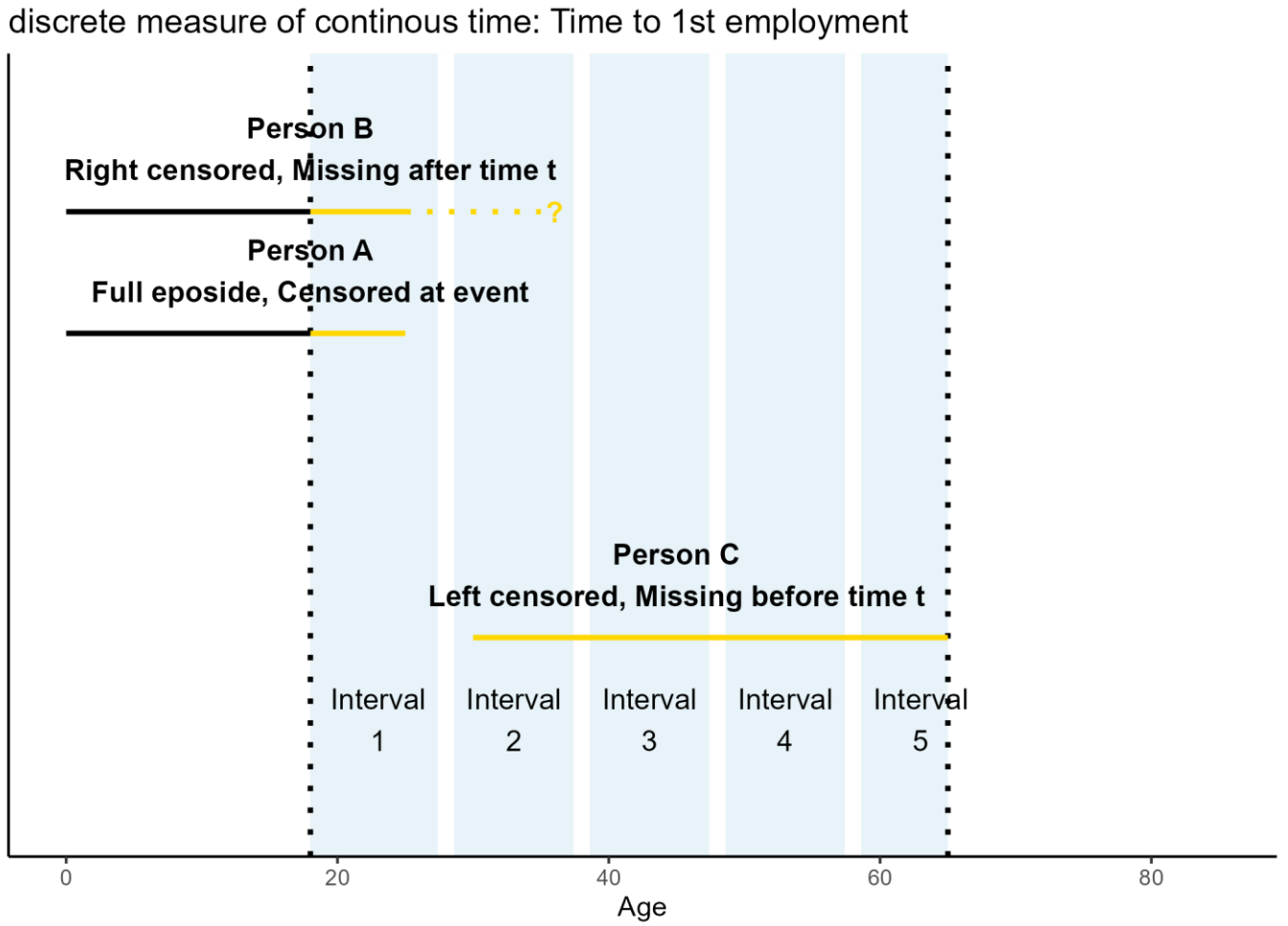
Censoring


Time to 1st employment
Person A
Full eposide, Censored at event
Start processtime
t18
End processtime
t65
0
20
40
60
80
Age

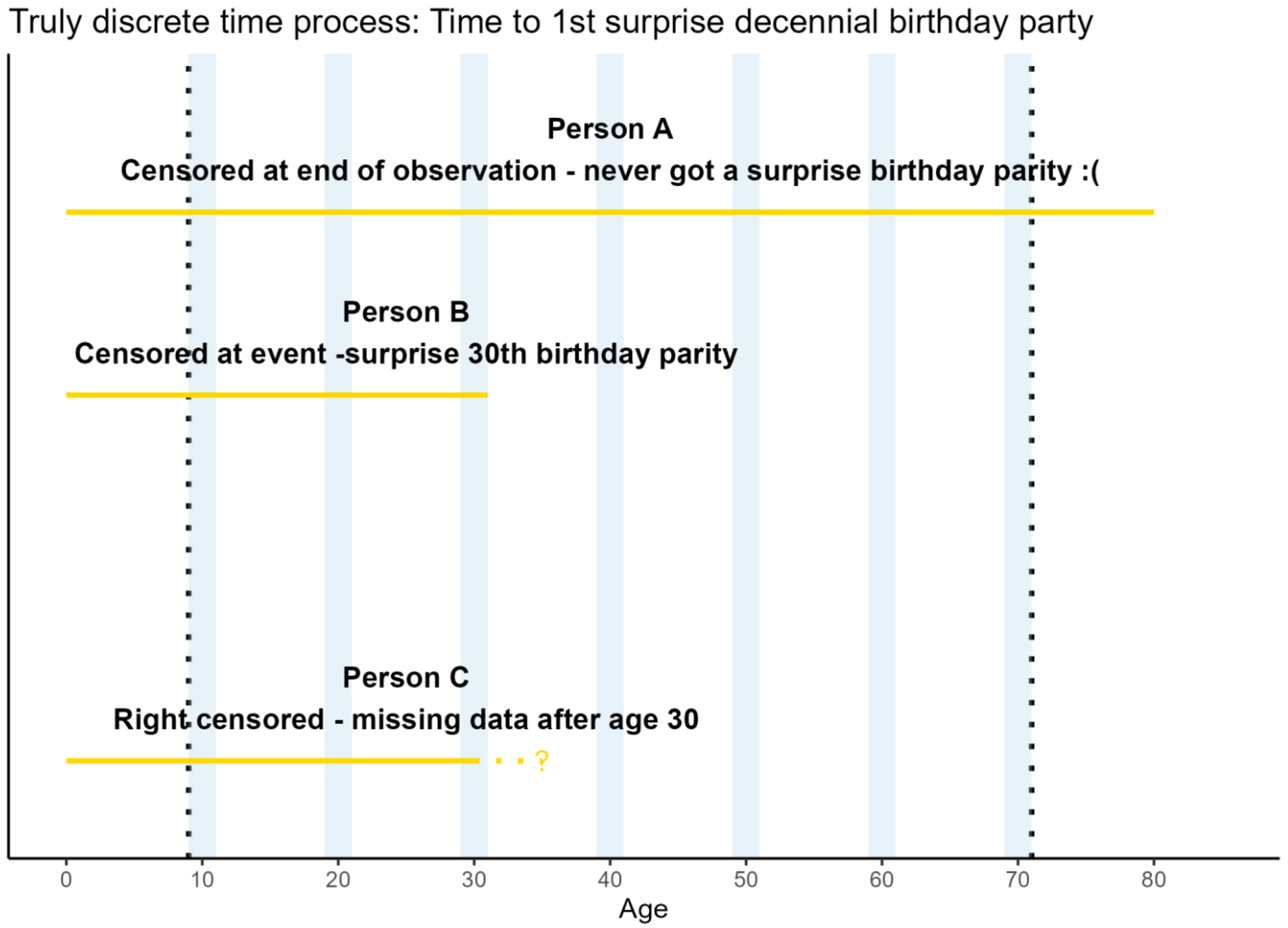
Censoring


Time to 1st employment
Person B
Right censored, Missing after time t
?
Person A
Full eposide, Censored at event
Start processtime
t18
End processtime
t65
Person C
Left censored, Missing before time t
0
20
40
60
80
Age

Censoring


discrete measure of continous time: Time to 1st employment
Person B
Right censored, Missing after time t
?
Person A
Full eposide, Censored at event
Person C
Left censored, Missing before time t
Interval 1
Interval 2
Interval 3
Interval 4
Interval 5
0
20
40
60
80
Age

Censoring


Truly discrete time process: Time to 1st surprise decennial birthday party
Person A
Censored at end of observation - never got a surprise birthday parity :(
Person B
Censored at event -surprise 30th birthday parity
Person C
Right censored - missing data after age 30
. . ?
0
10
20
30
40
50
60
70
80
Age

Recap - Continous and discrete time

- Choices and defintions required:
  - Model continous time process using continous EHA model ✓
  - Model continous time process using discrete EHA model ✓
  - Model discrete time process using discrete EHA model ✓
  - Model discrete time process using contious EHA model ✗

Key concepts of EHA

1. The What and Why of event history analysis ✓
2. The building blocks of time ✓
3. Censoring ✓
4. Continous vs discrete time ✓

Hand-ins / homework

- Think of study question and describe it using the concepts of todays lesson
- See detailed description in ”Hand-in nr 1.pdf” in Moodle
- Due Wednesday

Some literature for inspiration

	Event transition	Broader question / time process
Skopek et al 2020	PhD completion	How does Universities (EUI) influence PhD completion?
Sherkat 1991	Religious faith conversion	What factors predicts religious conversion over the life course?
Robinsson & Smit-Lovin 2001	Humor in Conversations	What is the structure of conversations? what predicts the use of humor in conversation
Ivanova et al 2013	Finding a partner after divorce	Does having a child prevent women from finding a partner after union separation?
Drefahl et al 2020	COVID mortality	What factors influence COVID mortality in Sweden?
De Mesquita & Smith 2010	Political Leaders duration in power	How do elected/non-elected leaders stay in power?
Blossfeld & Huinink 1991	Becoming a parent	Why do highly educated women have children later than less educated women?
Asadulla et al 2021	Having a son vs having a daughter	Are there still Son preferences an Bangladesh?
Andersson 2018	Fertility	Is there an effect of studying online for childbearing behavior?
Alsan & Goldin 2019	Child mortality	On the relationship between child mortality and installing clean water and sewage in 1800-1900 US

See Moodle for a complete literature reference list and links to all articles

That is it for today