Multi-state Modeling

The introduction and practice

Amarit Tansawet, MD, PhD.
15/11/2024 Sukho Place Building

Multi-state working group

C&B
Department of Clinical Epidemiology and Biostatistics

… and more to come

Recap

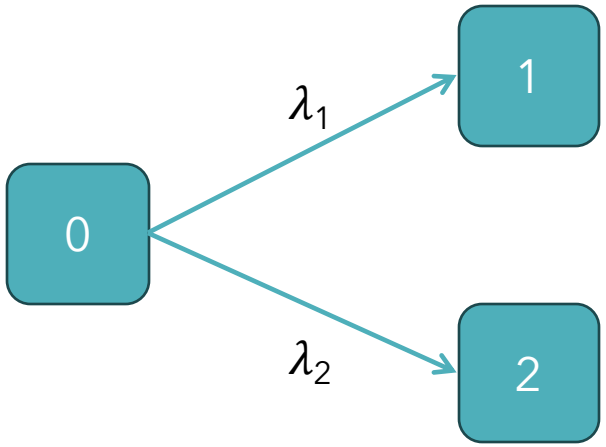
Hazard function

$$h(t) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Recap

Cause-specific hazard

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(t \leq T < t + \Delta t, E = k \,|\, T \geq t)}{\Delta t}$$



$$\sum_{e=1}^{K} \lambda_e(t) = \lim_{\Delta t \downarrow 0} \frac{\sum_{e=1}^{K} \mathrm{P}(t \leq T < t + \Delta t, E = e \,|\, T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(t \leq T < t + \Delta t \,|\, T \geq t)}{\Delta t} = h(t)\,.$$

Special case of multi-state model!

Multi-state model

Initial state

Intermediate state

Absorbing state

#0001

Bulbasaur

Grass · Poison

Transit 1

#0002

Ivysaur

Grass · Poison

Transit 2

#0003

Venusaur

Grass · Poison

Remain in the same state

Mutually exclusive

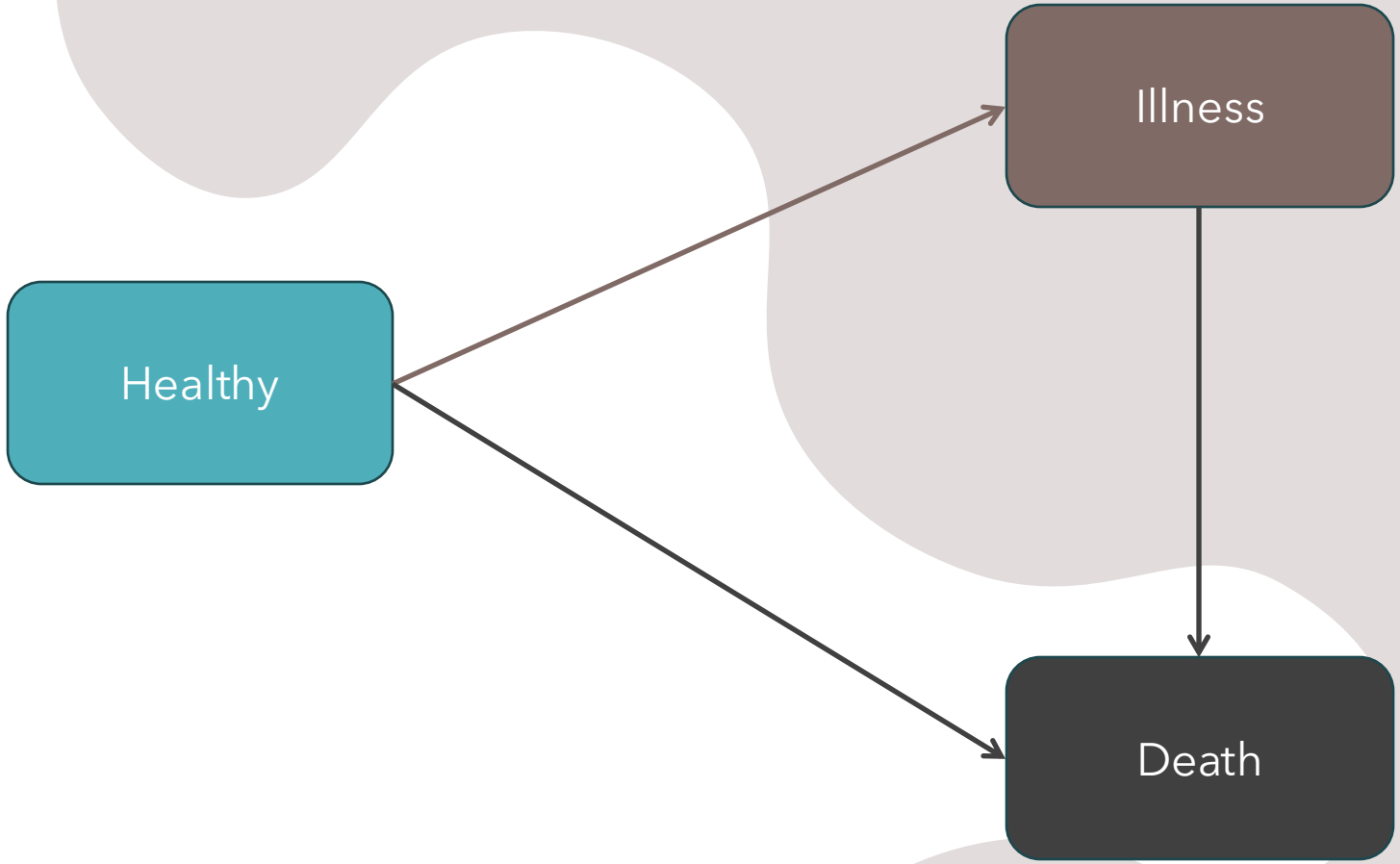
Why do we need multi-state model?

Need information on etiological mechanism

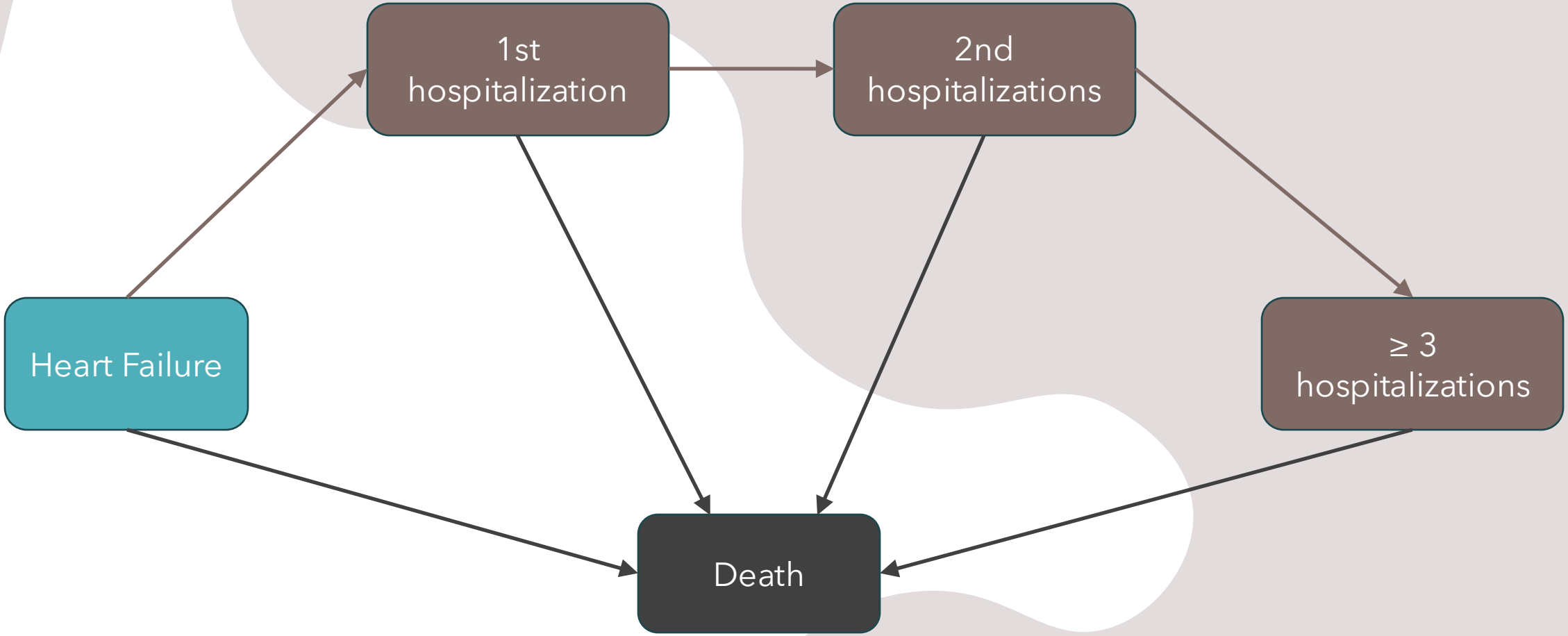
More accurate prediction

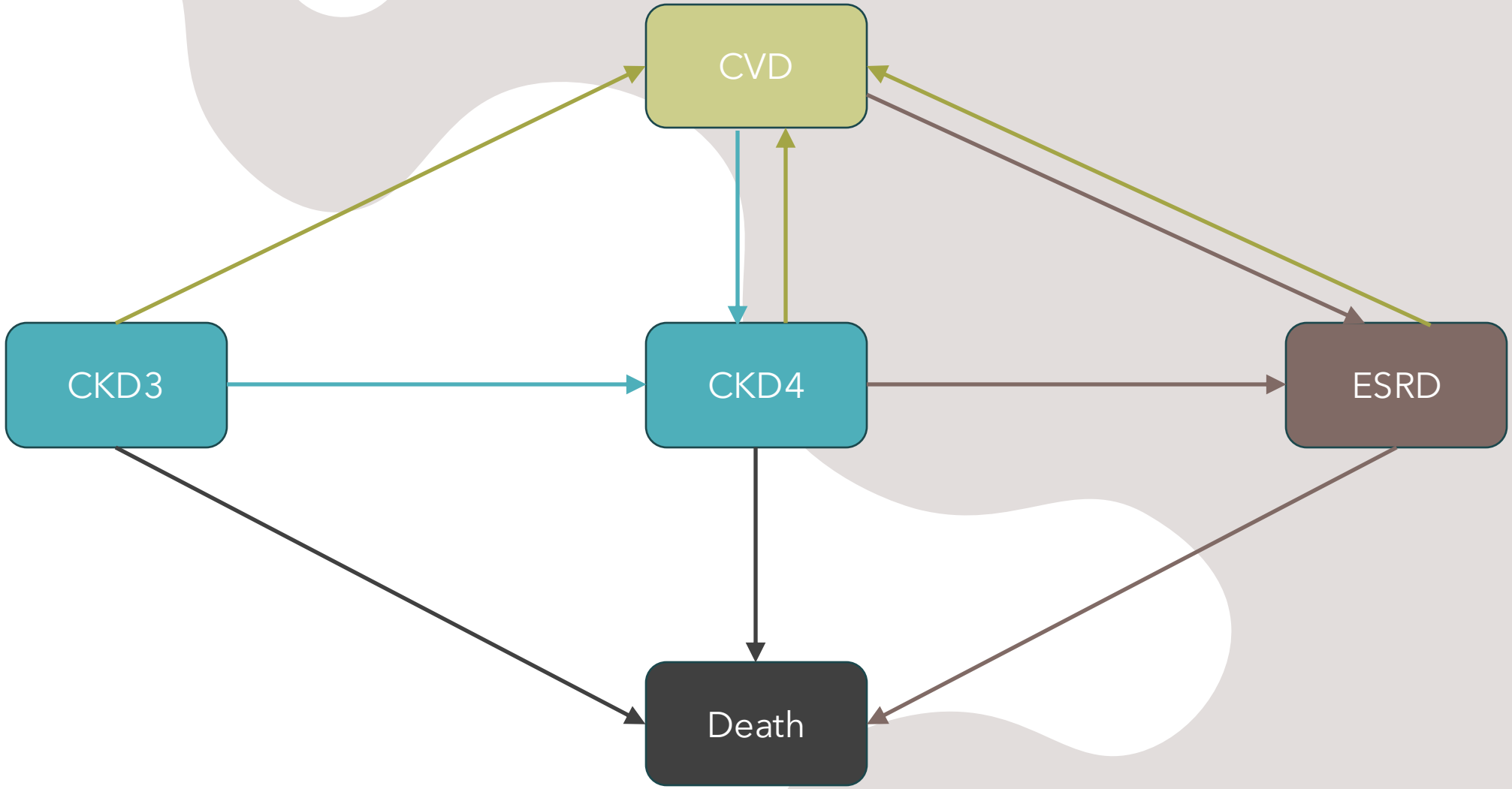
Patients don’t have to enter from the initial state

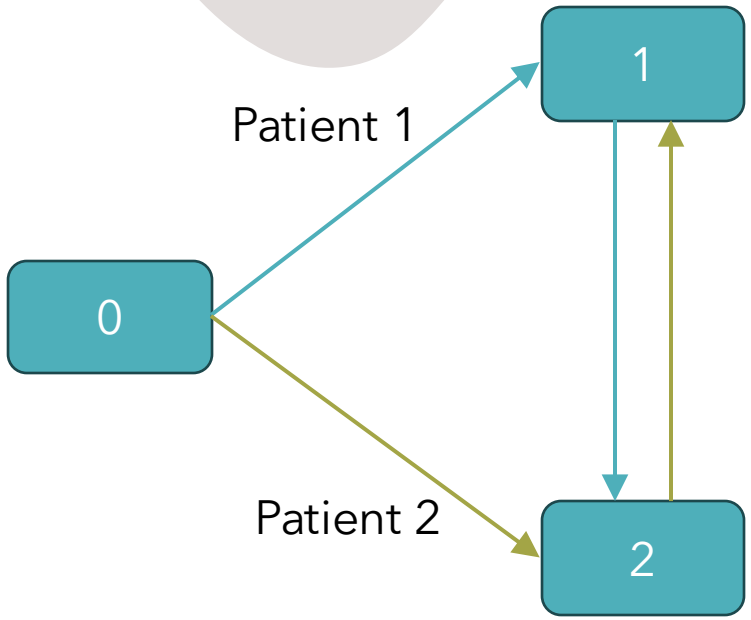
Illness-death model



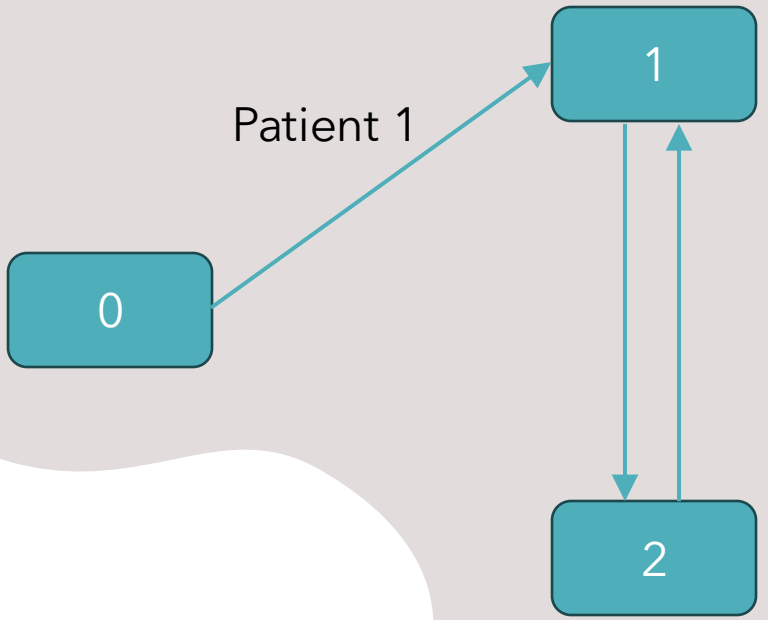
ACYCLIC


1st hospitalization
2nd hospitalizations
Heart Failure
≥ 3 hospitalizations
Death


CVD
CKD3
CKD4
ESRD
Death



?



Model as recurrent events?

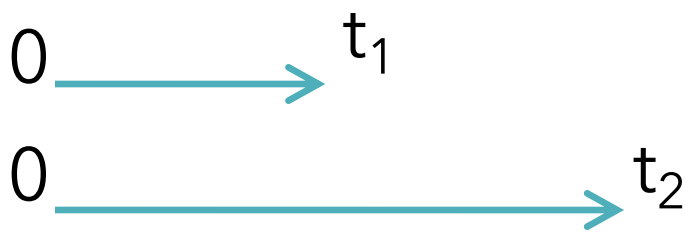
Markov property

What happens next only depends on the current state, not on what happened before

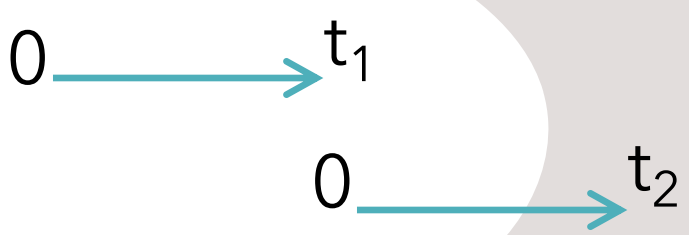
Need to be relaxed in more complex models

Time scale

Clock-forward



Clock-reset



Model different transitions with different time scales?

Basic data format: Stage-ordered format

ID	1.time	1.status	2.time	2.status	3.time	3.status
1	10	1	20	0	20	0
2	20	0	20	0	20	0
3	10	1	20	1	30	0
4	10	1	20	1	30	1
5	10	1	20	0	20	1
6	20	0	20	0	20	1

Basic data format

Stage-ordered format (irreversible multi-state data): One column contains the first observation time in that state and the other column is a zero-one variable that denotes whether the state was reached during follow-up

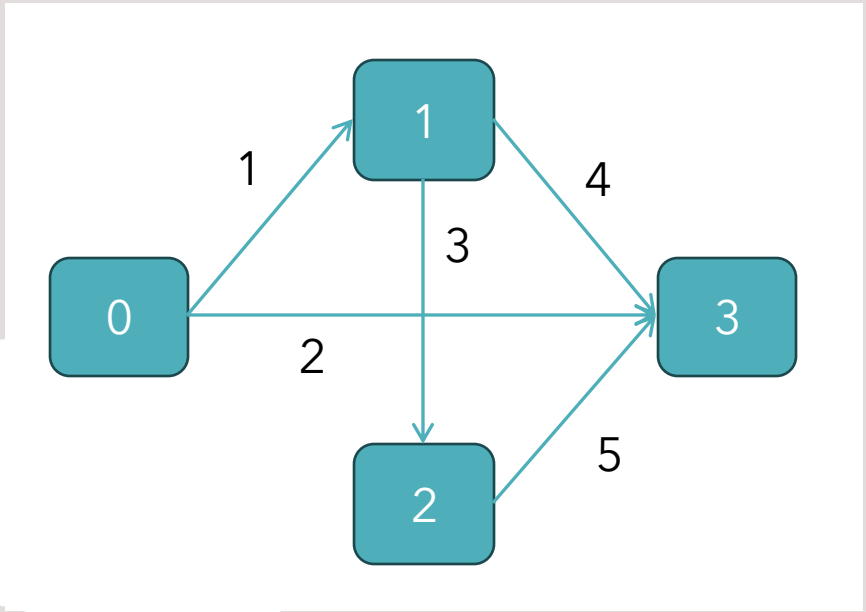
Basic data format: Stage-ordered format

ID	1.time	1.status	2.time	2.status	3.time	3.status
1	10	1	20	0	20	0
2	20	0	20	0	20	0
3	10	1	20	1	30	0
4	10	1	20	1	30	1
5	10	1	20	0	20	1
6	20	0	20	0	20	1

State diagram?

Basic data format: Stage-ordered format

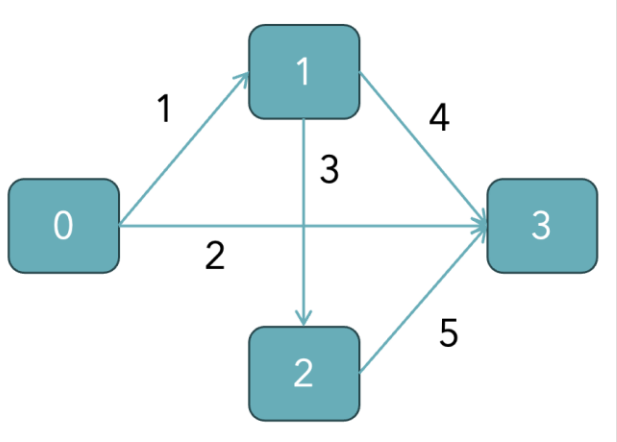
ID	1.time	1.status	2.time	2.status	3.time	3.status
1	10	1	20	0	20	0
2	20	0	20	0	20	0
3	10	1	20	1	30	0
4	10	1	20	1	30	1
5	10	1	20	0	20	1
6	20	0	20	0	20	1



Basic data format: Transition-based format

ID	1.time	1.status	2.time	2.status	3.time	3.status
1	10	1	20	0	20	0

ID	From	To	Trans	Start	Stop	Status
1	0	1	1	0	10	1
1	0	3	2	0	10	0
1	1	2	3	10	20	0
1	1	3	4	10	20	0

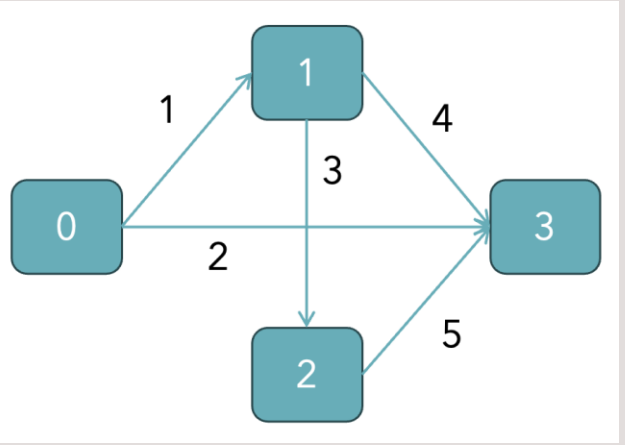


i.e., 1 → censor

Basic data format: Transition-based format

ID	1.time	1.status	2.time	2.status	3.time	3.status
2	20	0	20	0	20	0

ID	From	To	Trans	Start	Stop	Status
2	0	1	1	0	20	0
2	0	3	2	0	20	0

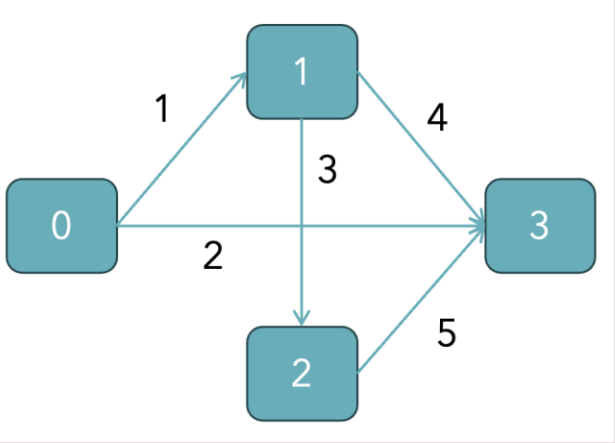


i.e., 0→ censor

Basic data format: Transition-based format

ID	1.time	1.status	2.time	2.status	3.time	3.status
3	10	1	20	1	30	0

ID	From	To	Trans	Start	Stop	Status
3	0	1	1	0	10	1
3	0	3	2	0	10	0
3	1	2	3	10	20	1
3	1	3	4	10	20	0
3	2	3	5	20	30	0

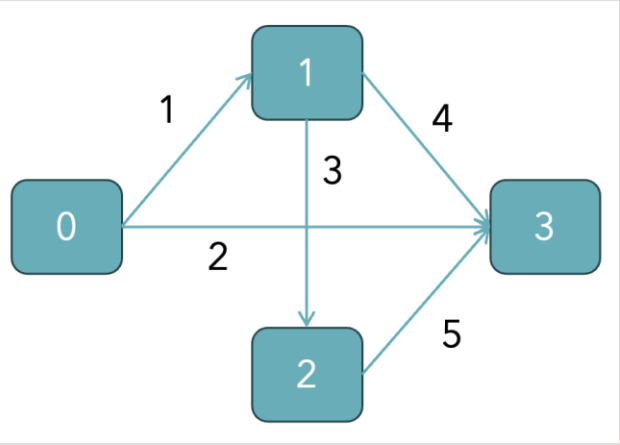


i.e., 2 → censor

Basic data format: Transition-based format

ID	1.time	1.status	2.time	2.status	3.time	3.status
4	10	1	20	1	30	1

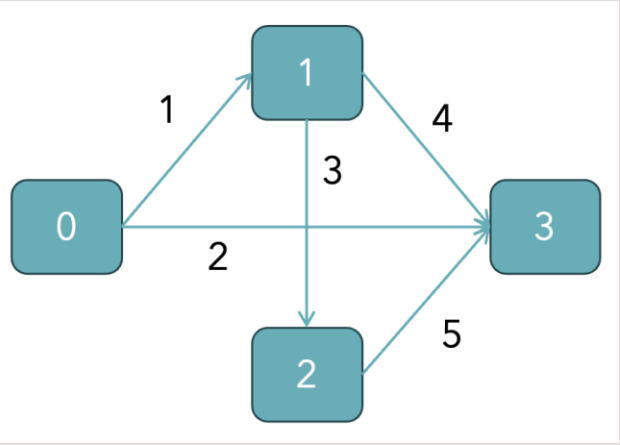
ID	From	To	Trans	Start	Stop	Status
4	0	1	1	0	10	1
4	0	3	2	0	10	0
4	1	2	3	10	20	1
4	1	3	4	10	20	0
4	2	3	5	20	30	1



Basic data format: Transition-based format

ID	1.time	1.status	2.time	2.status	3.time	3.status
5	10	1	20	0	20	1

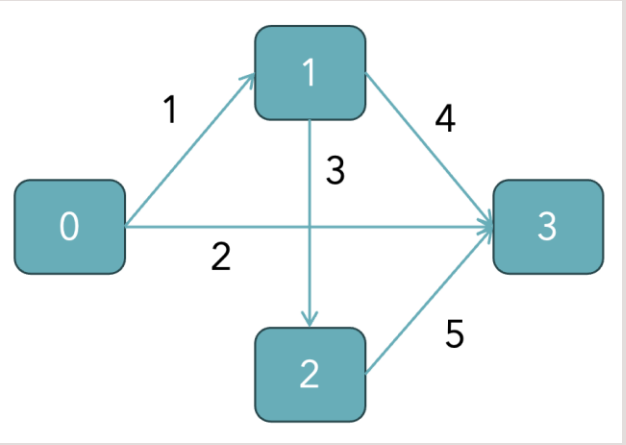
ID	From	To	Trans	Start	Stop	Status
5	0	1	1	0	10	1
5	0	3	2	0	10	0
5	1	2	3	10	20	0
5	1	3	4	10	20	1



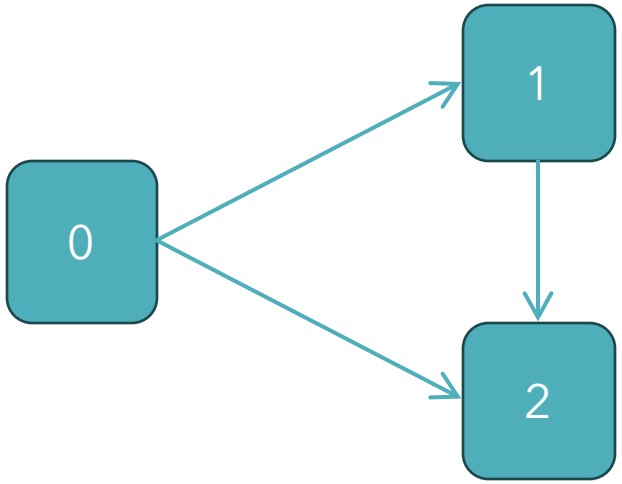
Basic data format: Transition-based format

ID	1.time	1.status	2.time	2.status	3.time	3.status
6	20	0	20	0	20	1

ID	From	To	Trans	Start	Stop	Status
6	0	1	1	0	20	0
6	0	3	2	0	20	1



Aalen-Johansen estimator



Irreversible illness-death model

$$\widehat{\mathrm{P}_{00}}(s,t) = \prod_{s<u\leq t} \left\{1 - \widehat{\lambda_{01}}(u) - \widehat{\lambda_{02}}(u)\right\}$$

$$\widehat{\mathrm{P}_{11}}(s,t) = \prod_{s<u\leq t} \left\{1 - \widehat{\lambda_{12}}(u)\right\}$$

$$\widehat{\mathrm{P}_{12}}(s,t) = \sum_{s<u\leq t} \widehat{\mathrm{P}_{11}}(s,u-)\widehat{\lambda_{12}}(u)$$

$$\widehat{\mathrm{P}_{01}}(s,t) = \sum_{s<u\leq t} \widehat{\mathrm{P}_{00}}(s,u-)\widehat{\lambda_{01}}(u)\widehat{\mathrm{P}_{11}}(u+,t)$$

$$\widehat{\mathrm{P}_{02}}(s,t) = \sum_{s<u<t} \widehat{\mathrm{P}_{00}}(s,u-)\widehat{\lambda_{01}}(u)\widehat{\mathrm{P}_{12}}(u+,t) + \sum_{s<u\leq t} \widehat{\mathrm{P}_{00}}(s,u-)\widehat{\lambda_{02}}(u)$$

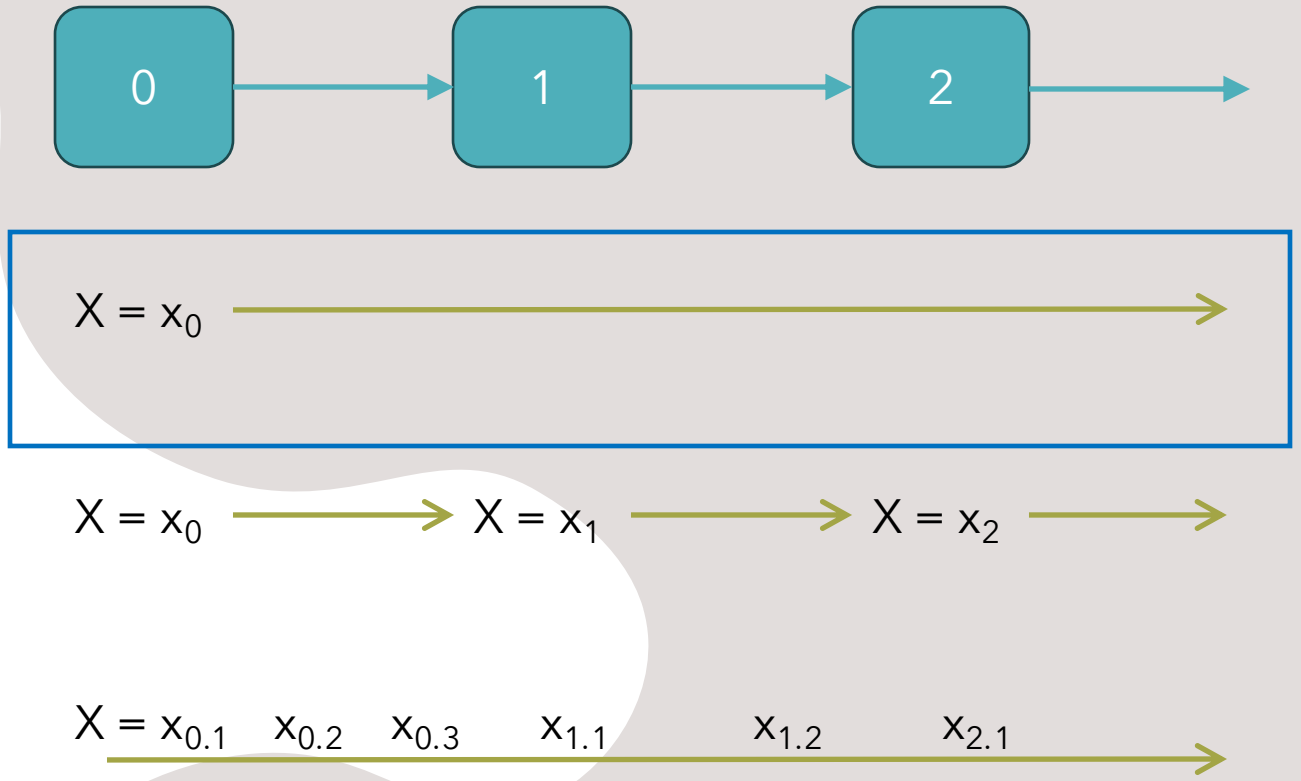
But if we would like to consider for the covariate pattern …

Covariate

Fixed at baseline

Transition specific

Time-varying within transition



Each transition has a different hazard function, which could have the same or different set of covariates

ID	From	To	Trans	Start	Stop	Status	HbA1c
3	0	1	1	0	10	1	7.2
3	0	3	2	0	10	0	7.2
3	1	2	3	10	20	1	7.2
3	1	3	4	10	20	0	7.2
3	2	3	5	20	30	0	7.2

ID	From	To	Trans	Start	Stop	Status	HbA1c
3	0	1	1	0	10	1	7.2
3	0	3	2	0	10	0	7
3	1	2	3	10	20	1	7.8
3	1	3	4	10	20	0	6.7
3	2	3	5	20	30	0	7

Modeling

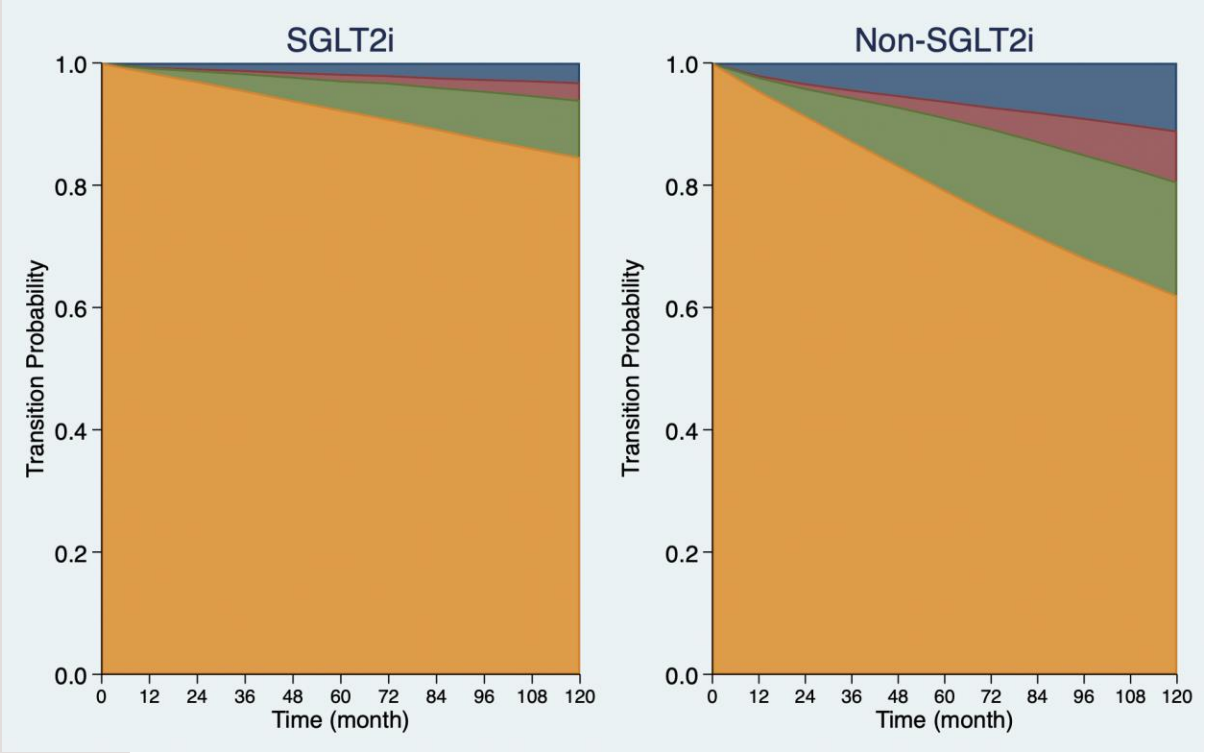
Cox PH model

Parametric survival model

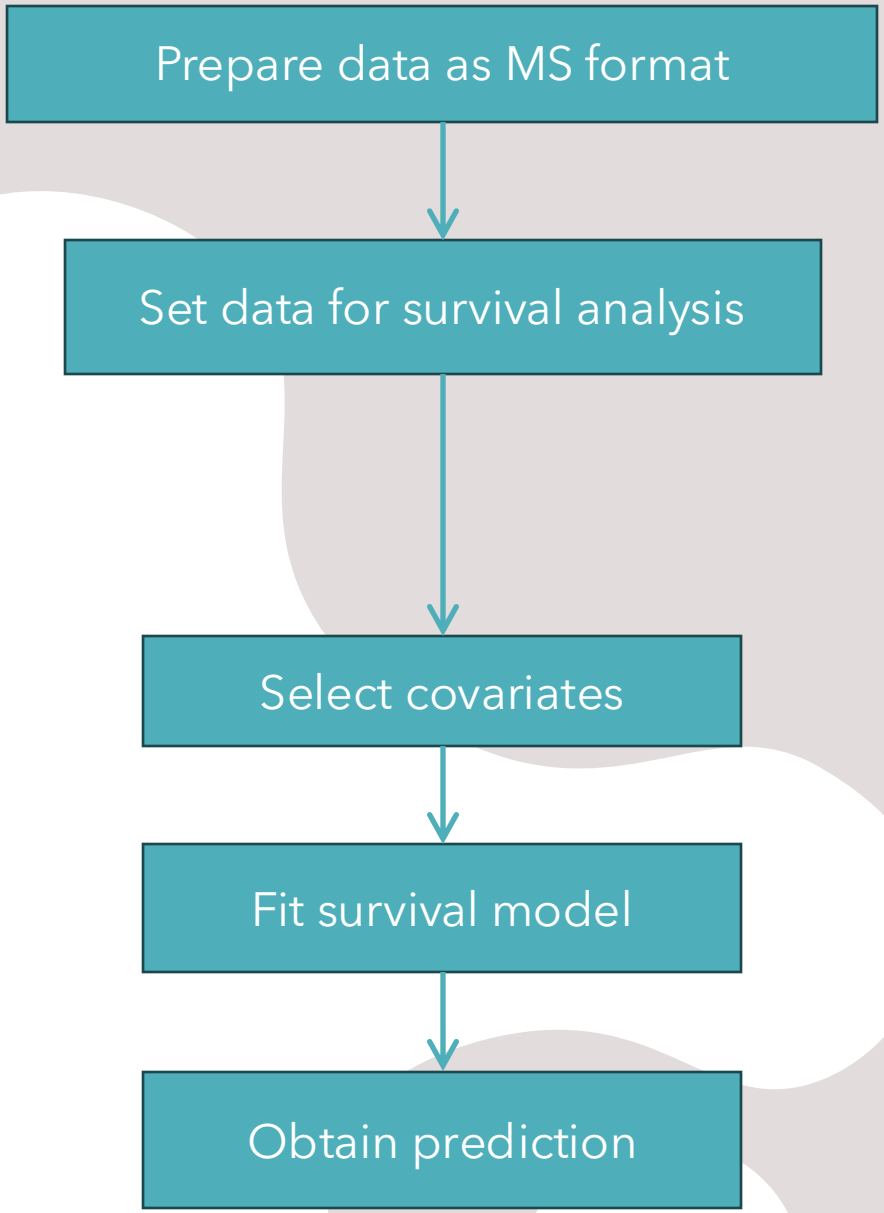
Spline-based model

Transition probability

the probability of a subject being in a certain stage given the subject's starting stage, the time at which the subject began moving through the process, and the subject's covariate profile.



Steps of analysis



Data for workshop

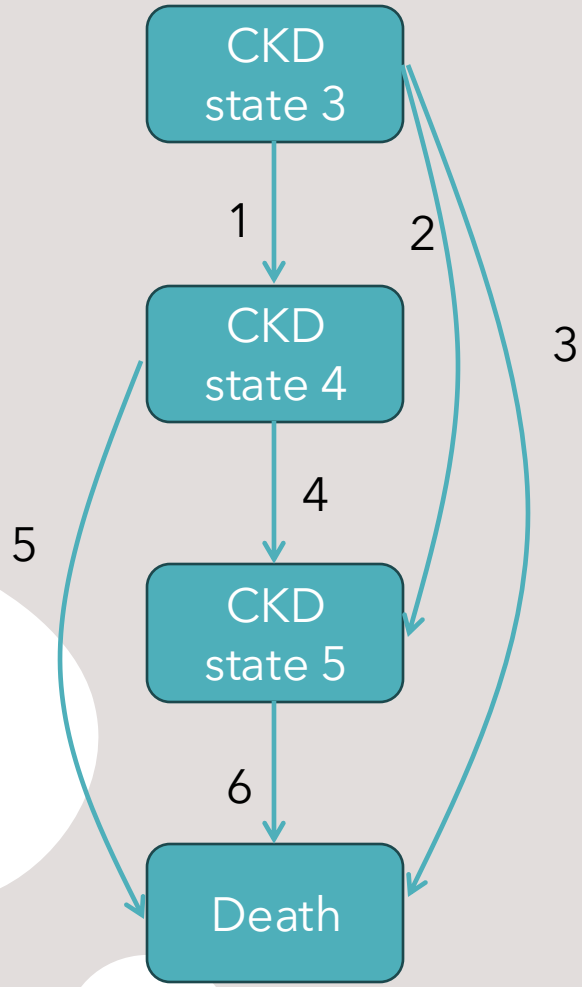
5266 patients

We want to know about the CKD progression

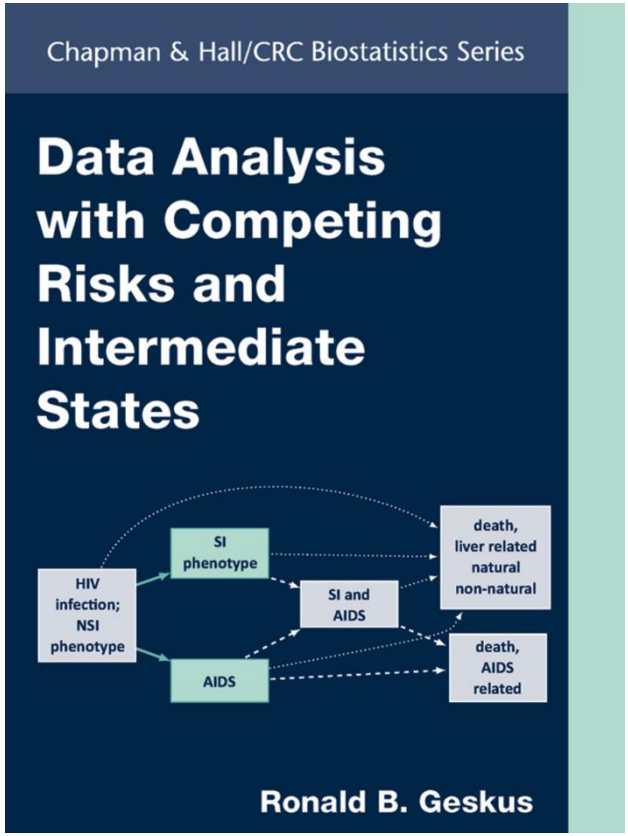
SGLT2i is the intervention of interest

(compared with other 2nd-medications)

Match 1:2 by the date of prescription



STATA time!



<< Suggested reading