

XVII Italian Stata Users Group Meeting



Modelling the risk of multimorbidity: an application of multistate models to the Swedish National March Cohort

Giulia Peveri^{1,2}, Rino Bellocco^{2,3}

¹ Dipartimento di Scienze Cliniche e di Comunità, Università degli Studi di Milano, Milano, Italia

² Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

³ Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca, Milano, Italia



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Karolinska
Institutet



Multimorbidity

- To address the coexistence of two or more diseases or conditions, the term **multimorbidity** is used
- Chronic diseases, defined as **health problems requiring ongoing management over a period of years or decades**¹ currently represent the predominant burden of healthcare²
- **Multistate models** are a well-suited statistical framework to address this problem

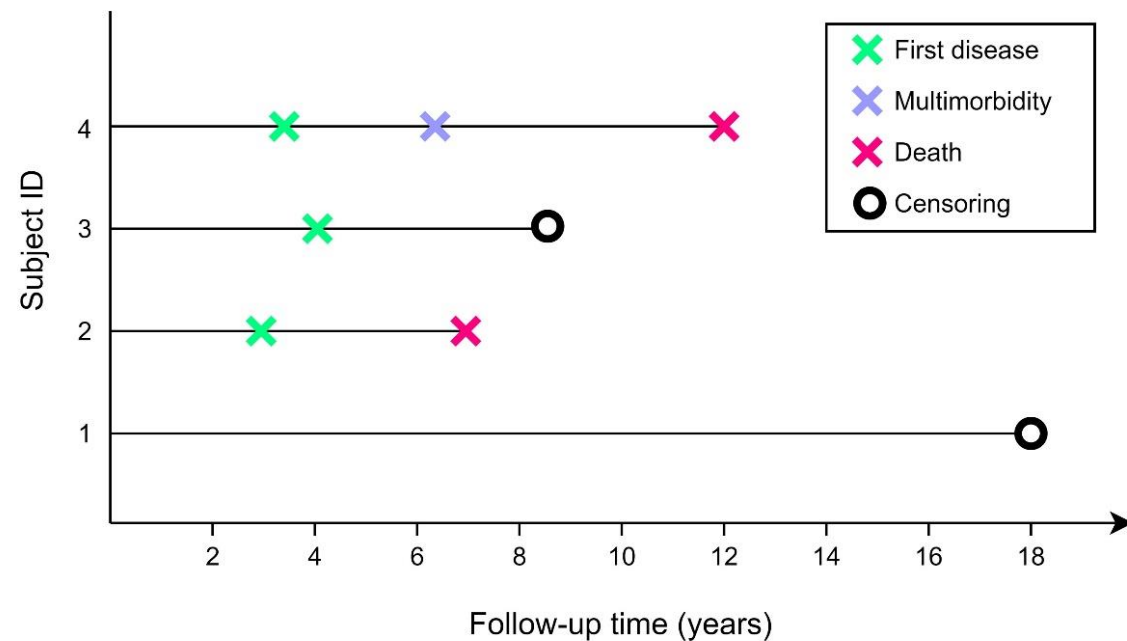
¹ World Health Organization, *Innovative care for chronic conditions: building blocks for action: global report*. Geneva: World Health Organization, 2002

² J. E. Bennett *et al.*, «NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4», *The Lancet*, 2018



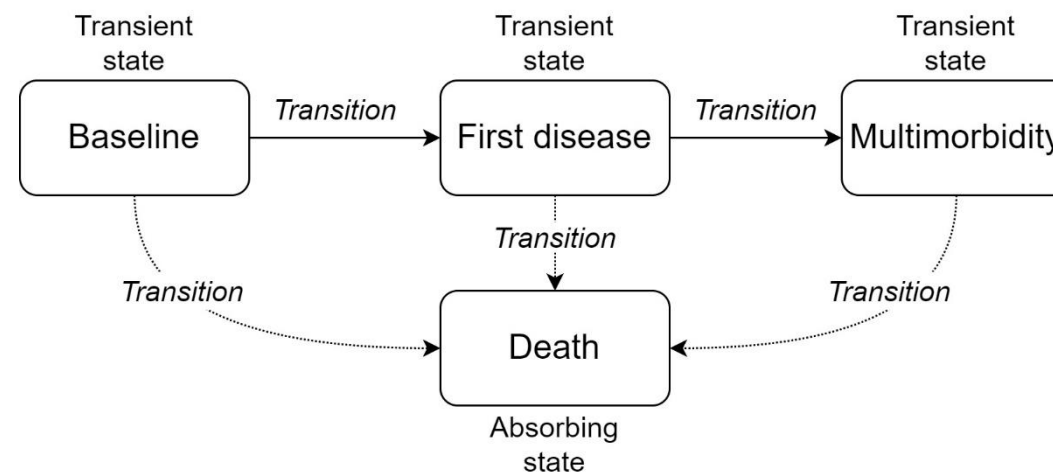
Multistate models

- In the setting of a longitudinal study, subjects are usually followed over time until the occurrence of a particular event of interest
- It is common to be interested in more than one event representing the patient's life experience



Multistate models

- The **states** can be defined on the basis of symptoms, clinical evaluations or biomarkers
- The events leading to a change of state are called **transitions**
- A state is *absorbing* if there cannot be further transitions from it. A state that is not absorbing is called *transient*
- Each transition between any two states is a (transition-specific) **survival model**

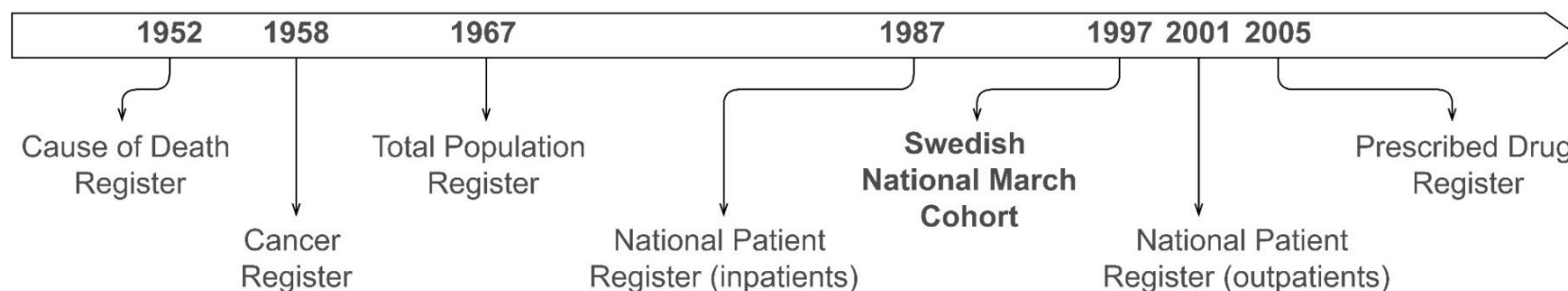


Aims

- Review of multistate modelling in **Stata** through an application using a large Swedish cohort
 - Data preparation
 - Choosing best fitting parametric model
 - Estimate **transition probabilities** between relevant diseases and **length of stay** in each state
 - Role of **covariates** on each transition

Study population

- The Swedish National March Cohort, established in 1997
- 3,600 cities across the country
- 43,865 participants³ linked to national registers



³ Y. Trolle Lagerros et al., «Cohort Profile: The Swedish National March Cohort», Int. J. Epidemiol, 2016

Outcomes⁴

- Multimorbidity: two of the following
 - **Cardiovascular diseases**, i.e., hypertensive heart disease (HHD), coronary heart disease (CHD), and stroke
 - **Selected cancers**, i.e., cancers of the gastrointestinal tract (colorectal, gastric, liver, pancreatic, and esophageal) and of the respiratory tract (lungs, bronchus, and trachea)
 - **Dementia**
 - **Chronic obstructive pulmonary disorder (COPD)**

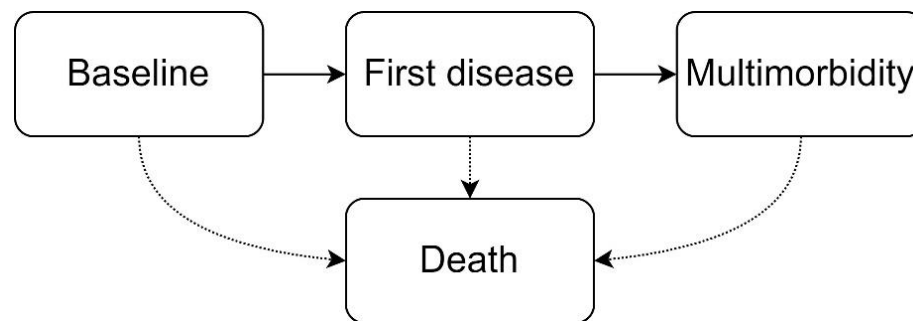
⁴ T. Vos et al., «Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019», The Lancet, 2020

Covariates

- Age, sex
- Lifestyle variables measured at baseline:
 - Cigarette smoking
 - Alcohol consumption
 - BMI
 - Physical activity
- Score from 0 (worst) to 4 (best) for each habit

Data structure

ID	age	sexf	smoke	alcohol	physact	bmi	st1	st2	st3	t1	t2	t3
1	76.3	M	4	4	2	1	0	0	0	19.2498	19.2498	19.2498
10	63.4	F	4	3	2	1	1	0	0	5.3279	19.2498	19.2498
48	58.5	F	1	1	4	3	1	0	1	13.5086	16.0219	16.0219
1209	51.7	M	4	0	4	2	1	1	0	18.6886	18.8939	19.2498
1297	67.8	M	4	2	2	1	1	1	1	12.4353	12.4435	13.6318



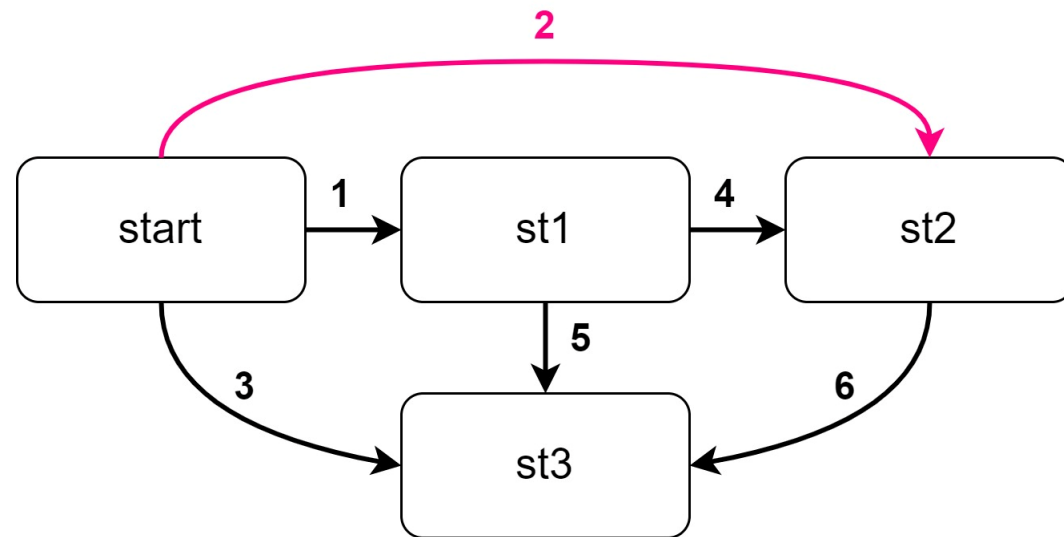
The msset command

```
. msset, id(ID) states(st1 st2 st3) times(t1 t2 t3)
```

```
. matrix list r(transmatrix)
```

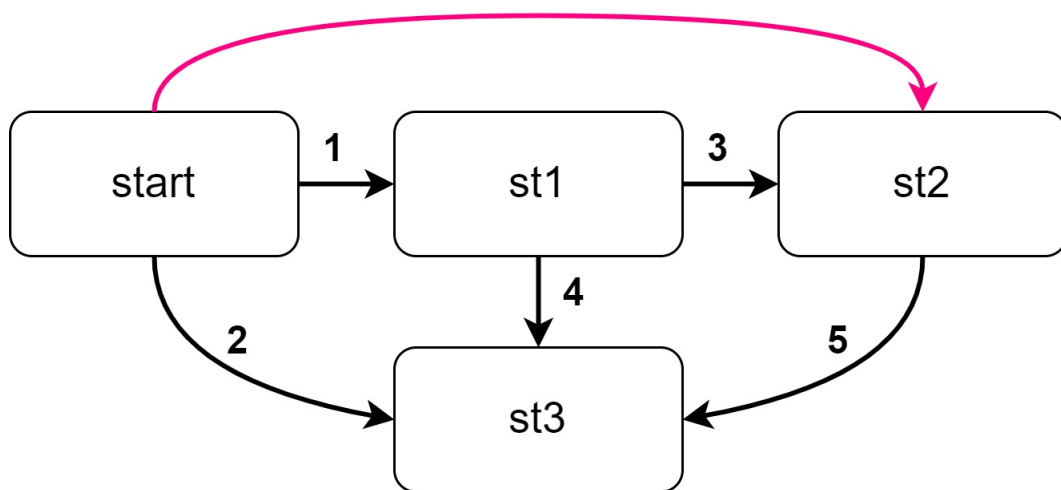
```
r(transmatrix)[4,4]
```

	to: start	to: st1	to: st2	to: st3
from:start	.	1	2	3
from:st1	.	.	4	5
from:st2	.	.	.	6
from:st3



The msset command

- Let's define a **transition matrix**



```

. mat tmat =
(.,1,.,2\.,.,3,4\.,.,.,5\.,.,.,.)

. mat colnames tmat = start st1 st2 st3

. mat rownames tmat = start st1 st2 st3

. mat list tmat
tmat[4,4]

```

	start	st1	st2	st3
start	.	1	.	2
st1	.	.	3	4
st2	.	.	.	5
st3

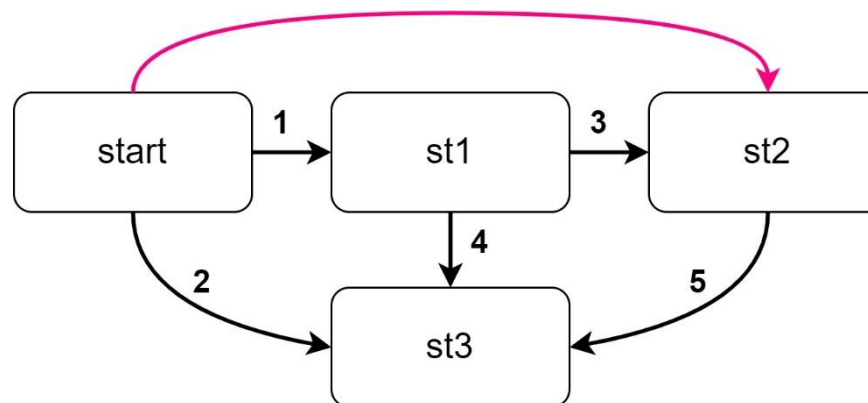
The msset command

```
. msset, id(ID) states(st1 st2 st3) times(t1 t2 t3) transmatrix(tmat)
```

```
. matrix list r(transmatrix)
```

```
r(transmatrix)[4,4]
```

	start	st1	st2	st3
start	.	1	.	2
st1	.	.	3	4
st2	.	.	.	5
st3

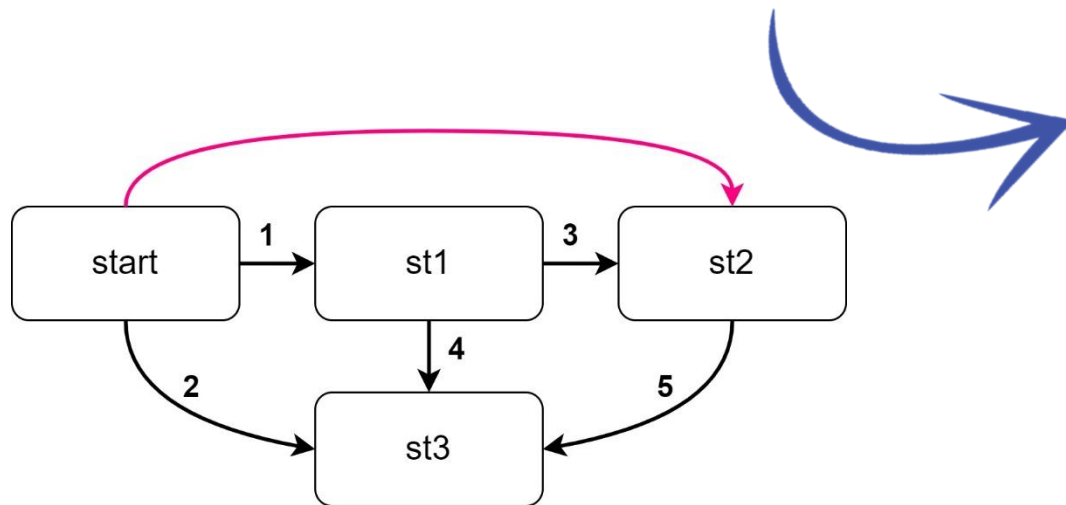


New data structure

ID	st1	st2	st3	t1	t2	t3
1	0	0	0	19.2498	19.2498	19.2498
48	1	0	1	13.5086	16.0219	16.0219
1297	1	1	1	12.4353	12.4435	13.6318

ID	_trans	_status	_start	_stop
1	1	0	0	19.249828
1	2	0	0	19.249828

ID	_trans	_status	_start	_stop
48	1	1	0	13.508555
48	2	0	0	13.508555
48	3	0	13.508555	16.021902
48	4	1	13.508555	16.021902



ID	_trans	_status	_start	_stop
1297	1	1	0	12.435318
1297	2	0	0	12.435318
1297	3	1	12.435318	12.443532
1297	4	0	12.435318	12.443532
1297	5	1	12.443532	13.631759

The stset command

```
. stset _stop, enter(_start) failure(_status=1)
```

Survival-time data settings

```

      Failure event: _status==1
Observed time interval: (0, _stop]
      Enter on or after: time _start
      Exit on or before: failure

```

```
-----
98,757 total observations
      0 exclusions
-----
```

```

98,757 observations remaining, representing
17,124 failures in single-record/single-failure data
1,419,884 total analysis time at risk and under observation
                                     At risk from t =          0
                                     Earliest observed entry t =          0
                                     Last observed exit t = 19.24983

```

Modelling transitions

- Now we can proceed modelling one transition at a time
- We fit different parametric models:
 - Exponential
 - Weibull
 - Royston-Parmar
 - Splines on the log-hazard scale
 - ...
- Then we choose the best fitting using AIC and/or BIC

Choosing the best fitting model

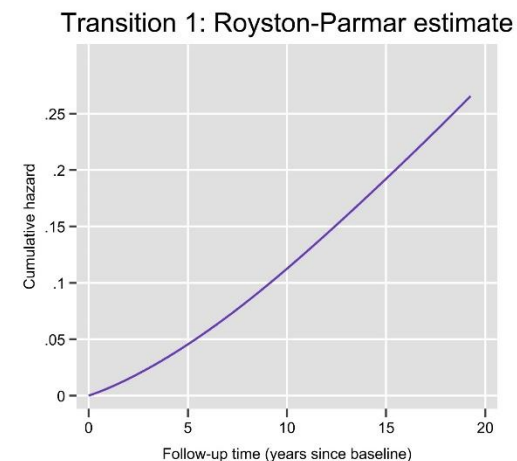
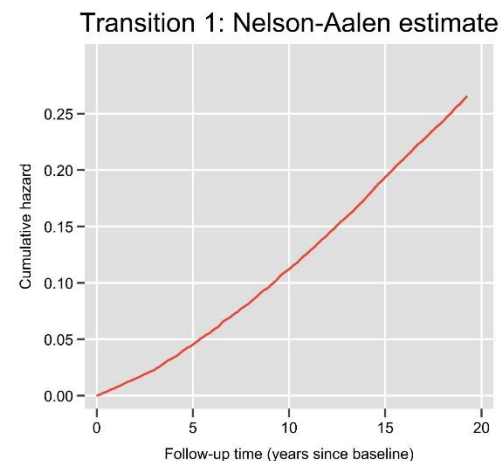
- Exponential/Weibull:


```
streg if _trans==1, distribution(exponential)
streg if _trans==1, distribution(weibull)
```
- Royston-Parmar


```
stpm2 if _trans==1, scale(hazard) df(3)
```
- Splines on the log-hazard scale


```
strcs if _trans==1, df(5) nohr
```
- To obtain AIC/BIC:


```
estat ic
```
- ...then the other transitions (RP, RP, Weibull, exponential)



The `stmerlin` command⁵

- `merlin` - mixed effects regression for linear, non-linear and user-defined models
- `stmerlin` - convenience wrapper for estimating a parametric and semi-parametric survival model with `merlin`

```
streg, distribution(exponential)
streg, distribution(weibull)
stpm2, scale(hazard) df(3)
strcs, df(5)
```



```
stmerlin, distribution(exponential)
stmerlin, distribution(weibull)
stmerlin, distribution(rp) df(3)
stmerlin, distribution(rcs) df(5)
```

⁵ <https://reddooranalytics.se/products/merlin/stmerlin/> by Michael Crowther, PhD

Checking PH assumption

```
. stmerlin age sexf smoke alcohol physact bmi if _trans==1, ///  
distribution(rp) df(3)
```

```
. stcox age sexf smoke alcohol physact bmi if _trans==1  
. estat phtest, detail rank
```

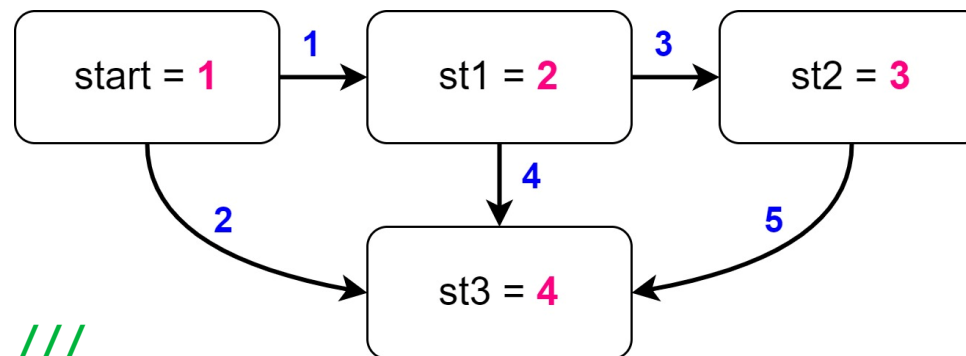
- We find that H_0 is rejected for sex and physact (in a Cox model)

```
. stmerlin age sexf smoke alcohol physact bmi if _trans==1, ///  
distribution(rp) df(3) tvc(sexf physact) dftvc(1)  
. estimates store m1
```

- ...then the other transitions

The predictms command

- . range time 0 20 5 // first last nobs
- . predictms, transmatrix(tmat) los prob ci ///
at1(age 60) models(m1 m2 m3 m4 m5) timevar(time)
- . predictms, from(2) transmatrix(tmat) los prob ci ///
at1(age 60) models(m1 m2 m3 m4 m5) timevar(time)
- . predictms, from(3) transmatrix(tmat) los prob ci ///
at1(age 60) models(m1 m2 m3 m4 m5) timevar(time)



The predictms command

- `_los_at1_a_b`: length of stay in state b for `at1(...)`, given they started from state `a`

```
. predictms, transmatrix(tmat) los from(1) at1(age 60) models(m1 m2 m3 m4 m5) timevar(time)
```

```
. list _los_at1_* in 1/5, clean
```

time	<code>_los_at1_1_1</code>	<code>_los_at1_1_2</code>	<code>_los_at1_1_3</code>	<code>_los_at1_1_4</code>
0	0	0	0	0
5	4.8476443	.10925481	.00410224	.03899866
10	9.2561739	.48645762	.04357779	.21379074
15	13.086128	1.1234589	.16822656	.62218696
20	16.219033	1.9198402	.41435584	1.4467711

↑
Baseline

↑
First disease

↑
Multimorbidity

↑
Death

← The sum is equal to time

The predictms command

- `_los_at1_a_b`: length of stay in state b for `at1(...)`, given they started from state `a`

```
. predictms, transmatrix(tmat) los from(2) at1(age 60) models(m1 m2 m3 m4 m5) timevar(time)
```

```
. list _los_at1_* in 1/5, clean
```

time	_los_at1_2_1	_los_at1_2_2	_los_at1_2_3	_los_at1_2_4
0	0	0	0	0
5	0	4.1373179	.16338393	.69929818
10	0	7.0991726	.70512803	2.1956993
15	0	9.0965389	1.4812265	4.4222345
20	0	10.312432	2.3001926	7.3873751

↑
First disease

↑
Multimorbidity

↑
Death

← The sum is equal to time

The predictms command

- `_los_at1_a_b`: length of stay in state b for `at1(...)`, given they started from state `a`

```
. predictms, transmatrix(tmat) los from(3) at1(age 60) models(m1 m2 m3 m4 m5) timevar(time)
```

```
. list _los_at1_* in 1/5, clean
```

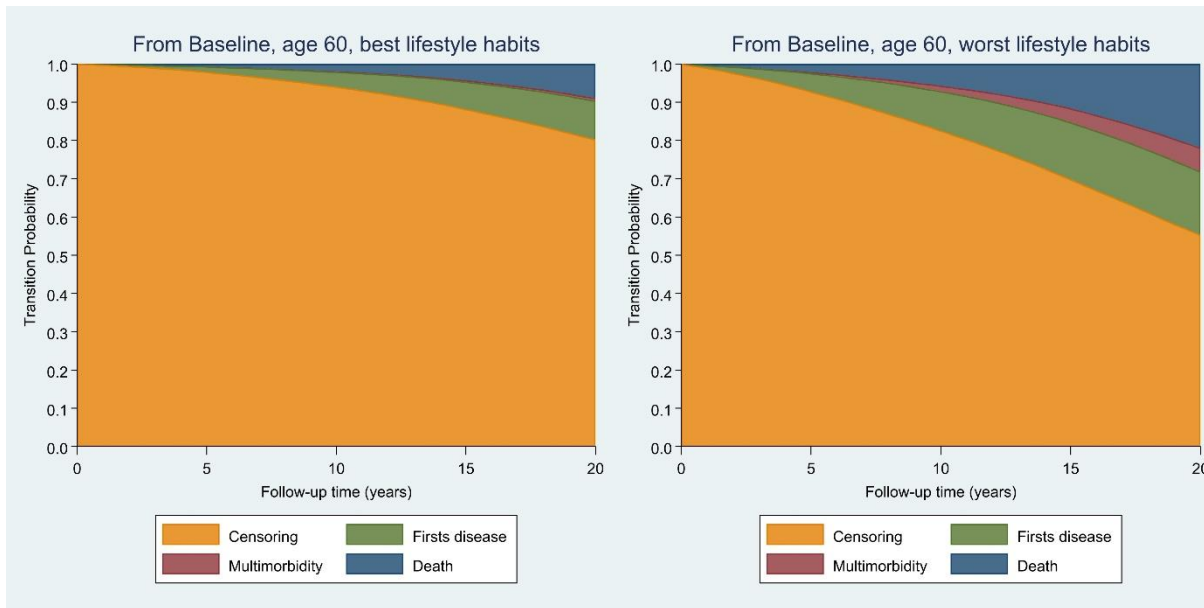
time	<code>_los_at1_3_1</code>	<code>_los_at1_3_2</code>	<code>_los_at1_3_3</code>	<code>_los_at1_3_4</code>
0	0	0	0	0
5	0	0	3.8859586	1.1140414
10	0	0	6.1897502	3.8102498
15	0	0	7.5569233	7.4430767
20	0	0	8.3693825	11.630618

↑
Multimorbidity

↑
Death

← The sum is equal to time

Transition probabilities using graphms



- ```

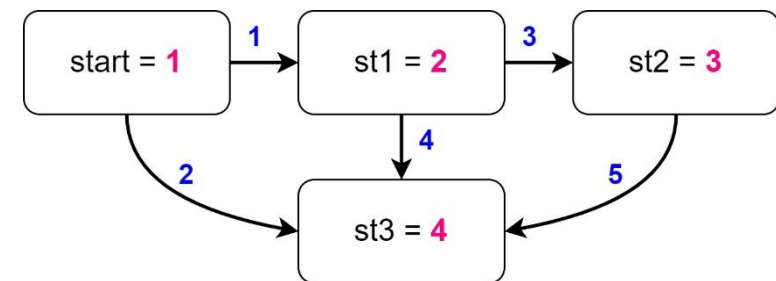
predictms, transmatrix(tmat) prob ///
models(m1 m2 m3 m4 m5) at1(age 60 ///
smoke 4 alcohol 4 physact 4 bmi 4) ///
timevar(time)

```
- ```

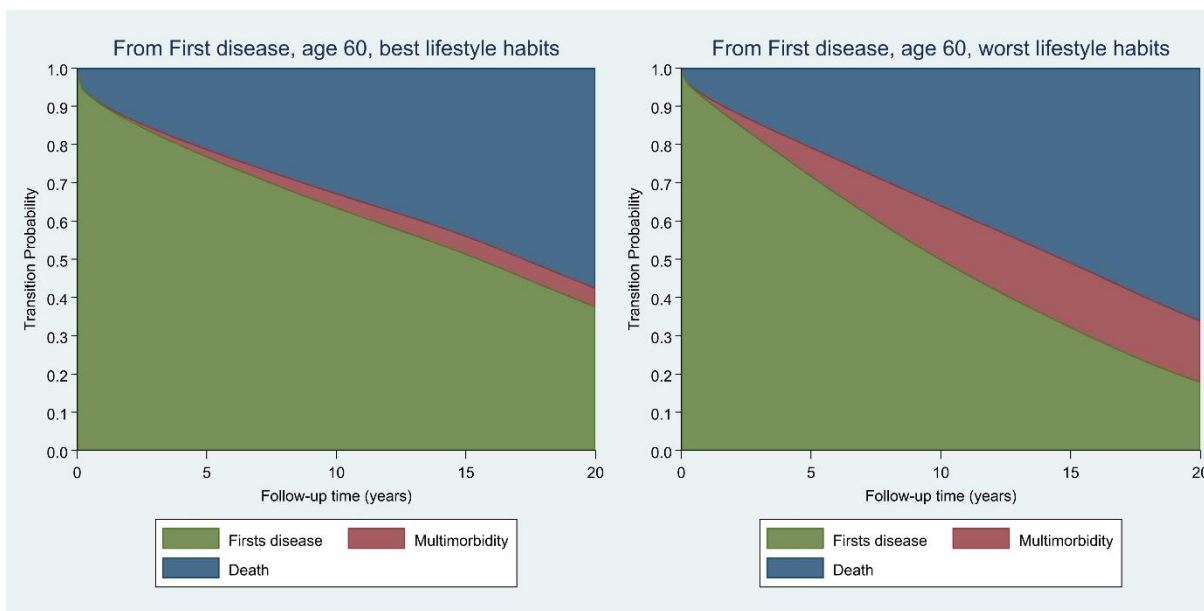
graphms, from(1) nstates(4) timevar(time)

```

(and then same code with smoke 0 alcohol 0 physact 0 bmi 0 for worst lifestyle)



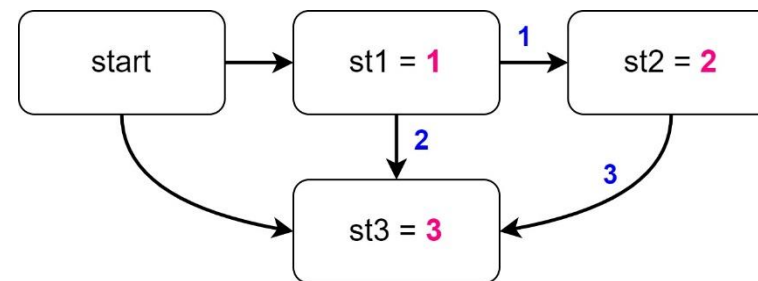
Transition probabilities using graphms



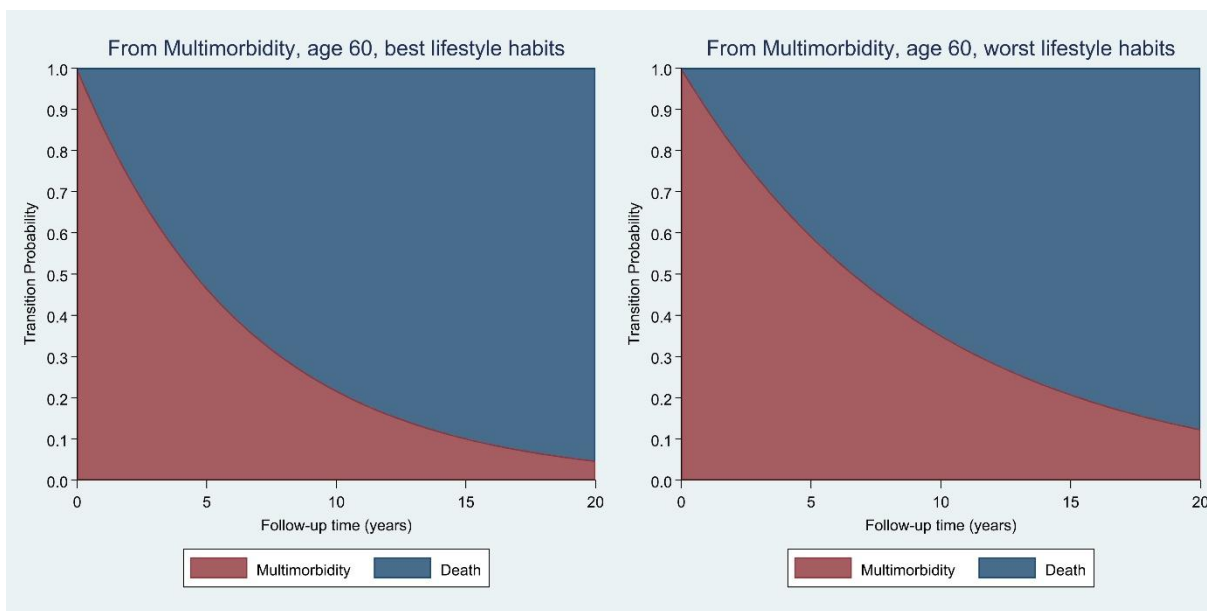
- . mat tmat1 = (.,1,2\.,.,3\.,.,.)
- . mat colnames tmat1 = st1 st2 st3
- . mat rownames tmat1 = st1 st2 st3

- . predictms, transmatrix(tmat1) prob ///
- models(m3 m4 m5) at1(age 60 ///
- smoke 4 alcohol 4 physact 4 bmi 4) ///
- timevar(time)

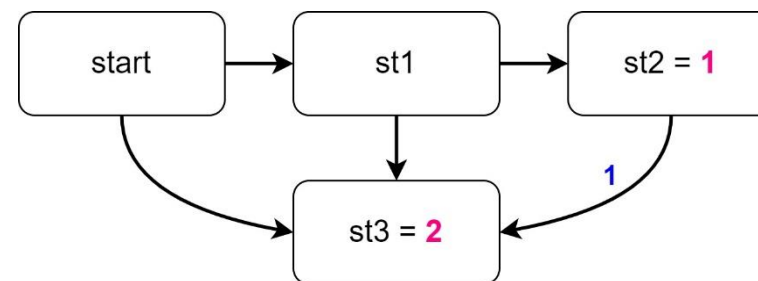
- . graphms, from(1) nstates(3) timevar(time)



Transition probabilities using graphms



- . mat tmat2 = (.,1\.,.)
- . mat colnames tmat2 = st2 st3
- . mat rownames tmat1 = st2 st3
- . predictms, transmatrix(tmat2) ///
 prob models(m5) at1(age 60 ///
 smoke 4 alcohol 4 physact 4 bmi 4) ///
 timevar(time)
- . graphms, from(1) nstates(2) timevar(time)



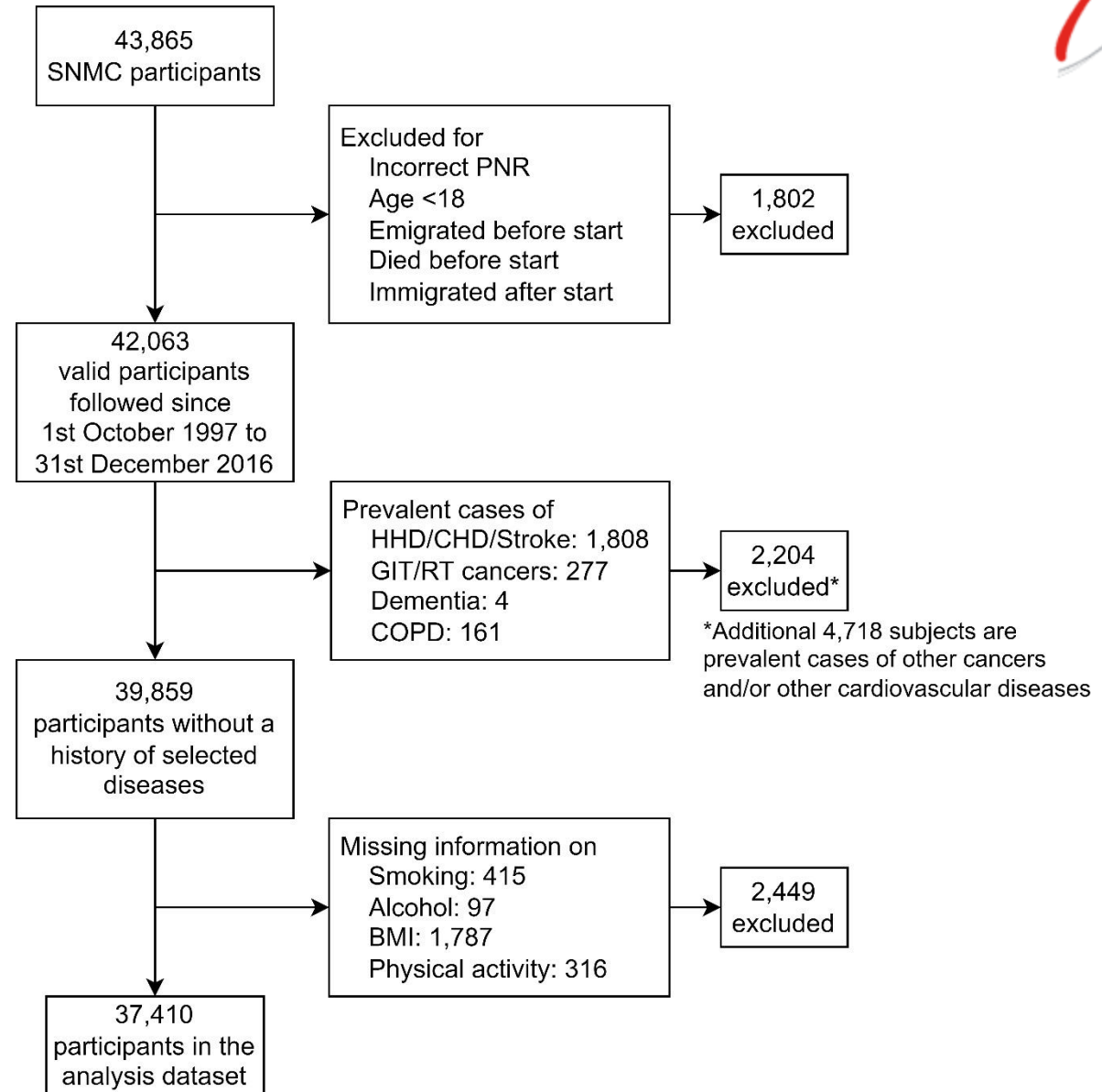
- Joint work with:
 - Rino Bellocco^{1,2}
 - Weimin Ye¹
 - Ylva Trolle Lagerros¹
 - Vincenzo Bagnardi²
 - Antonella Zambon²

Thank you!

¹ Karolinska Institutet, Stockholm

² Università degli Studi di Milano-Bicocca, Milano

Cohort selection



Transitions

