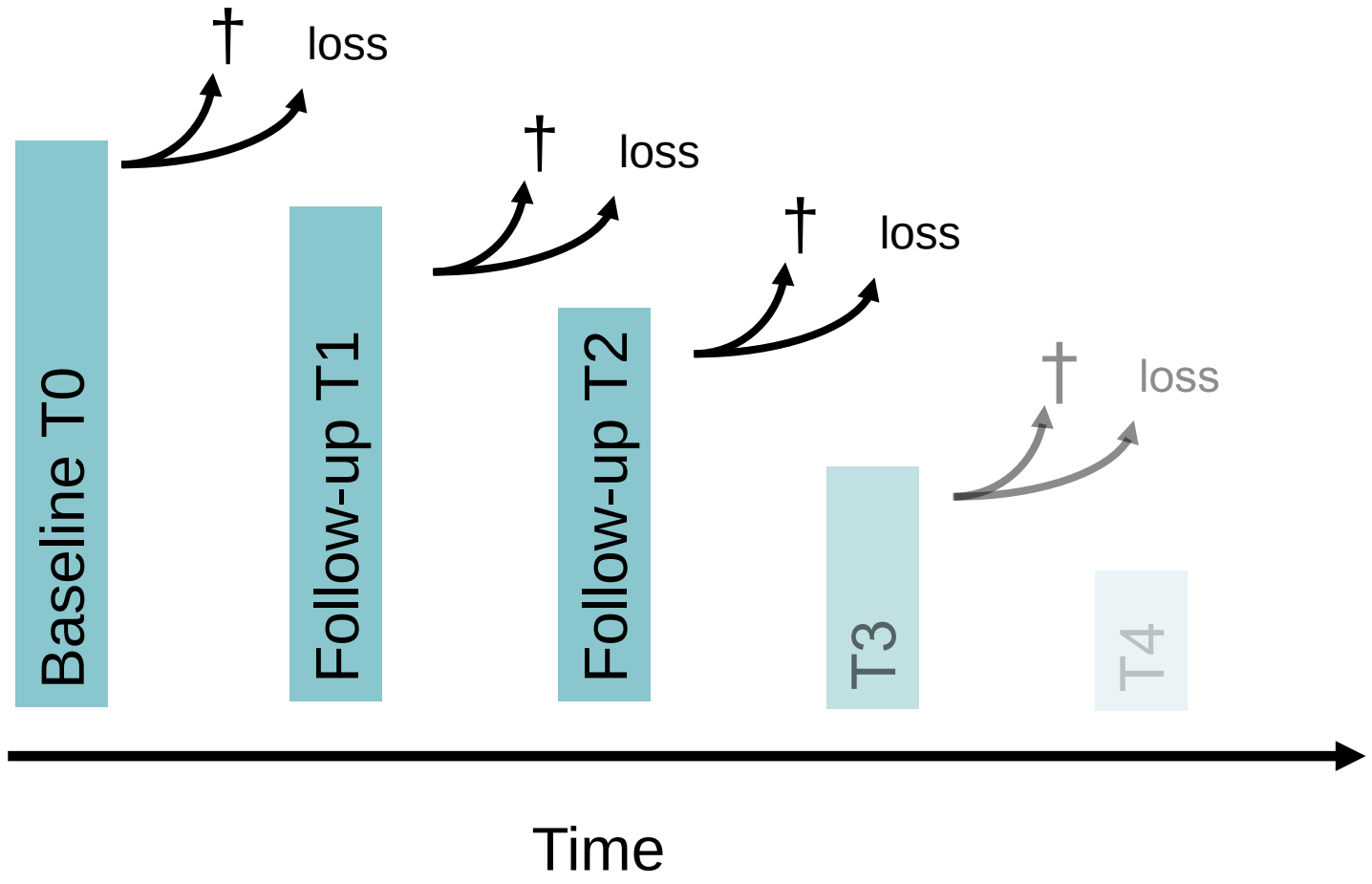
ulm university universität ulm

Raphael Peter
Institute of Epidemiology and Medical Biometry

Investigating trajectories in observational studies using joint latent class mixture time-to-event models

Common situation in longitudinal observational studies

- Scores
  (Depression, QoL, …)
- Biomarker
- Sensor data



Outline

1. Proof of Concept - Example using generated data

2. Depression trajectories and risk of cardiovascular events — KAROLA study
   - competing events

3. Physical activity trajectories and mortality — ActiFE Ulm study
   - covariates in longitudinal sub-model
   - alternative time scale
   - special link function

4. Pros and Cons

Software used

Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm

Cécile Proust-Lima
Inserm U1219
Université de Bordeaux

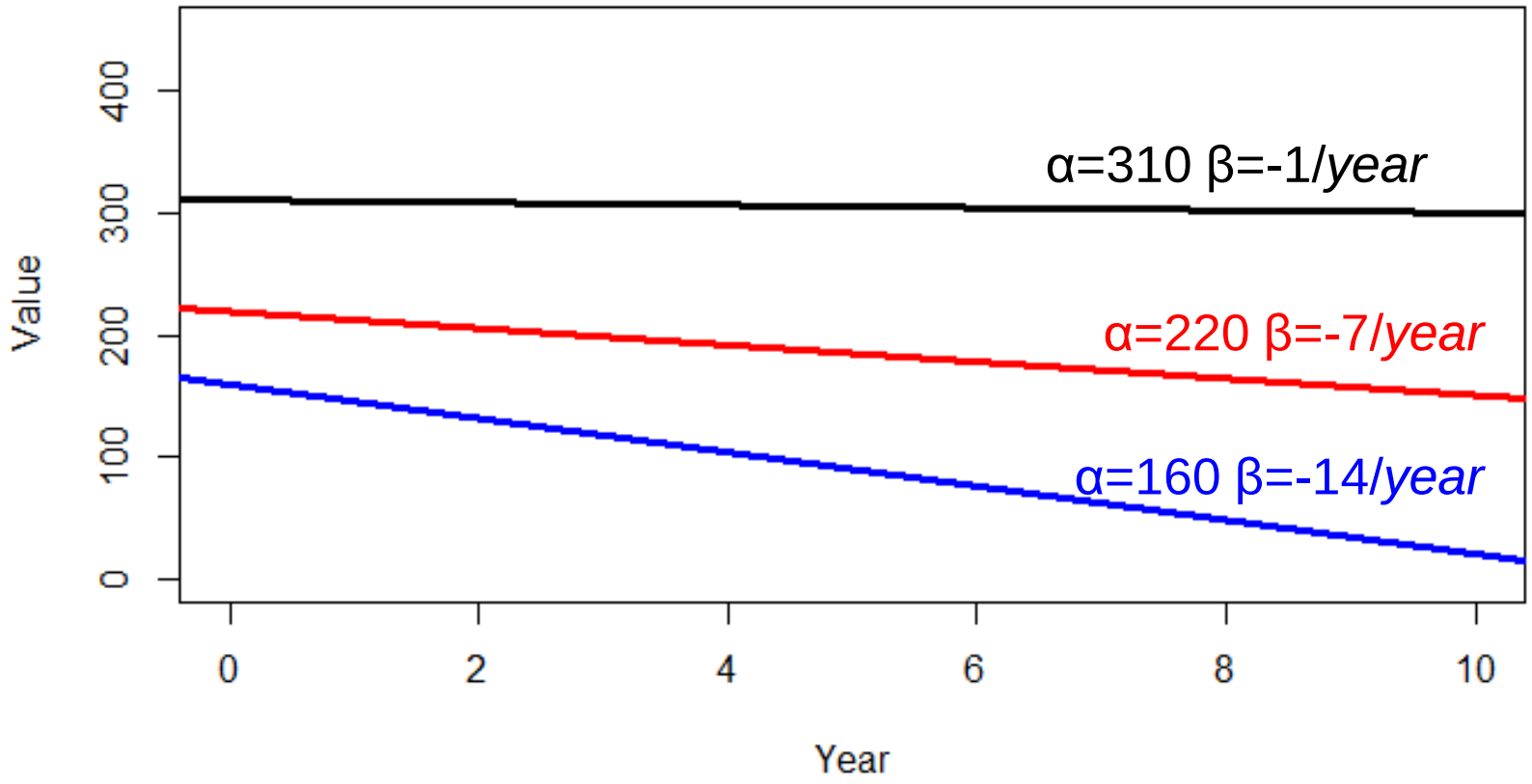
Viviane Philipps
Inserm U1219
Université de Bordeaux

Benoit Liquet
LMAP, UMR CNRS 5142
Université de Pau
et des Pays de L'Adour

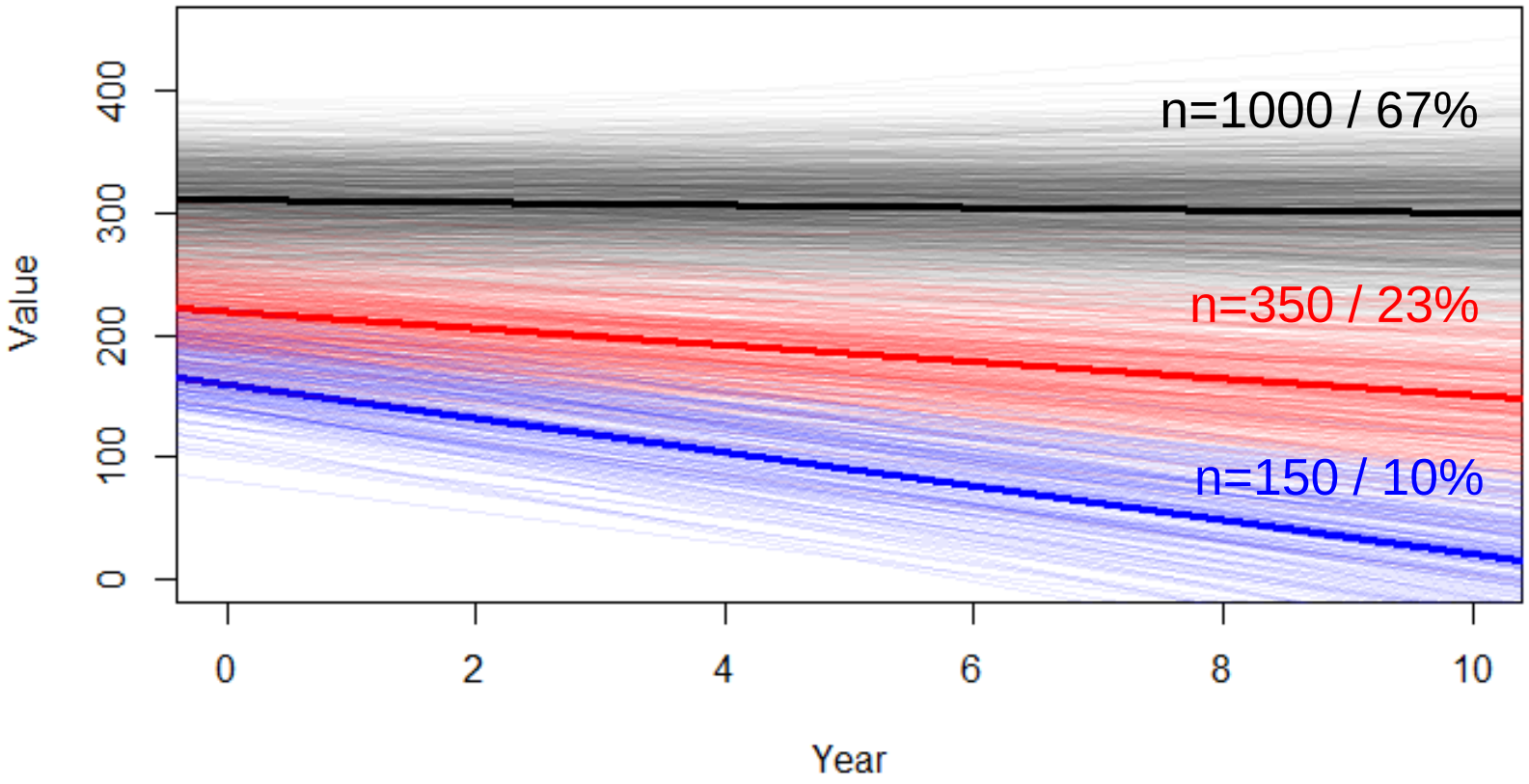
doi:10.18637/jss.v078.i02

1. Proof of Concept

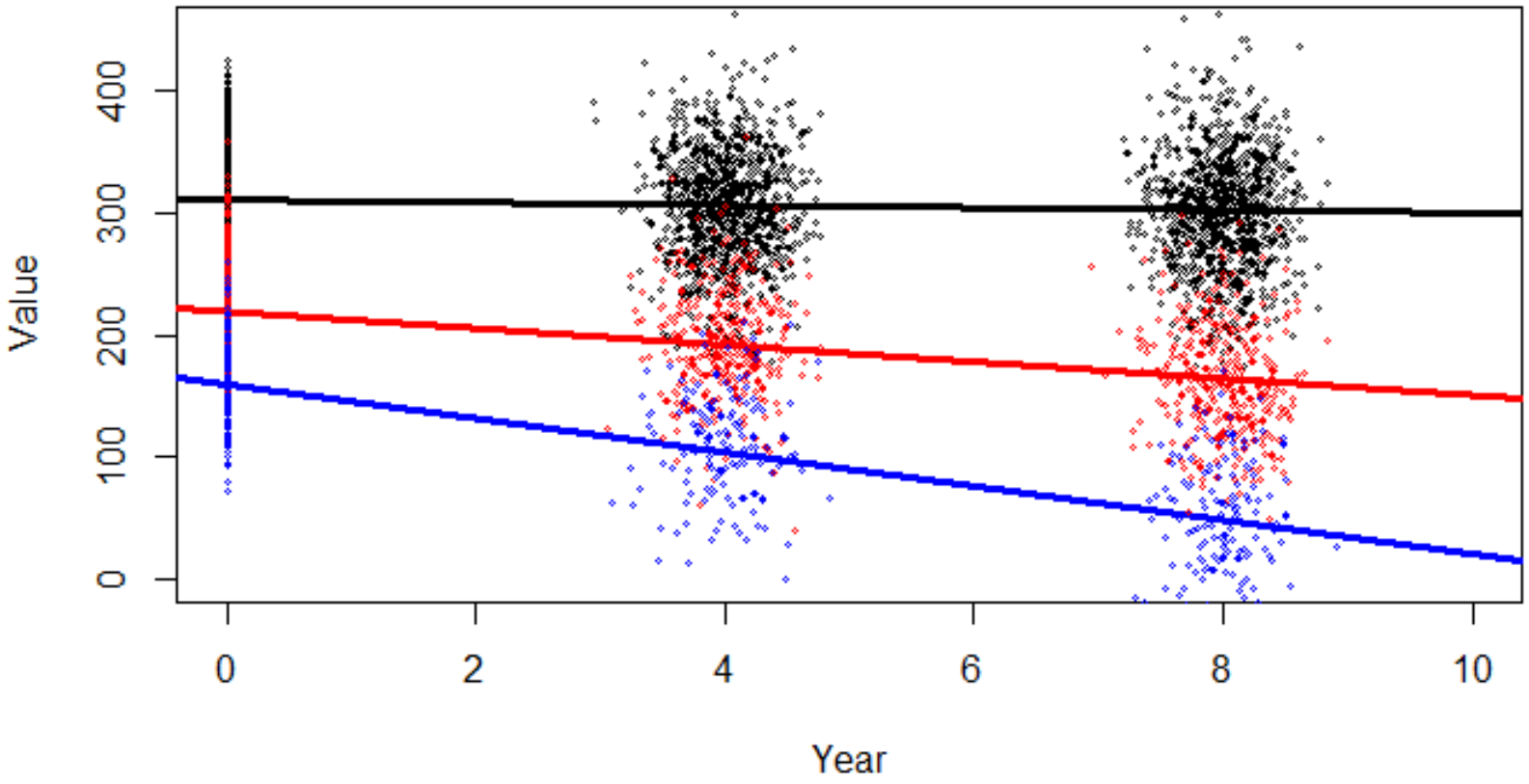
Assumed trajectory classes



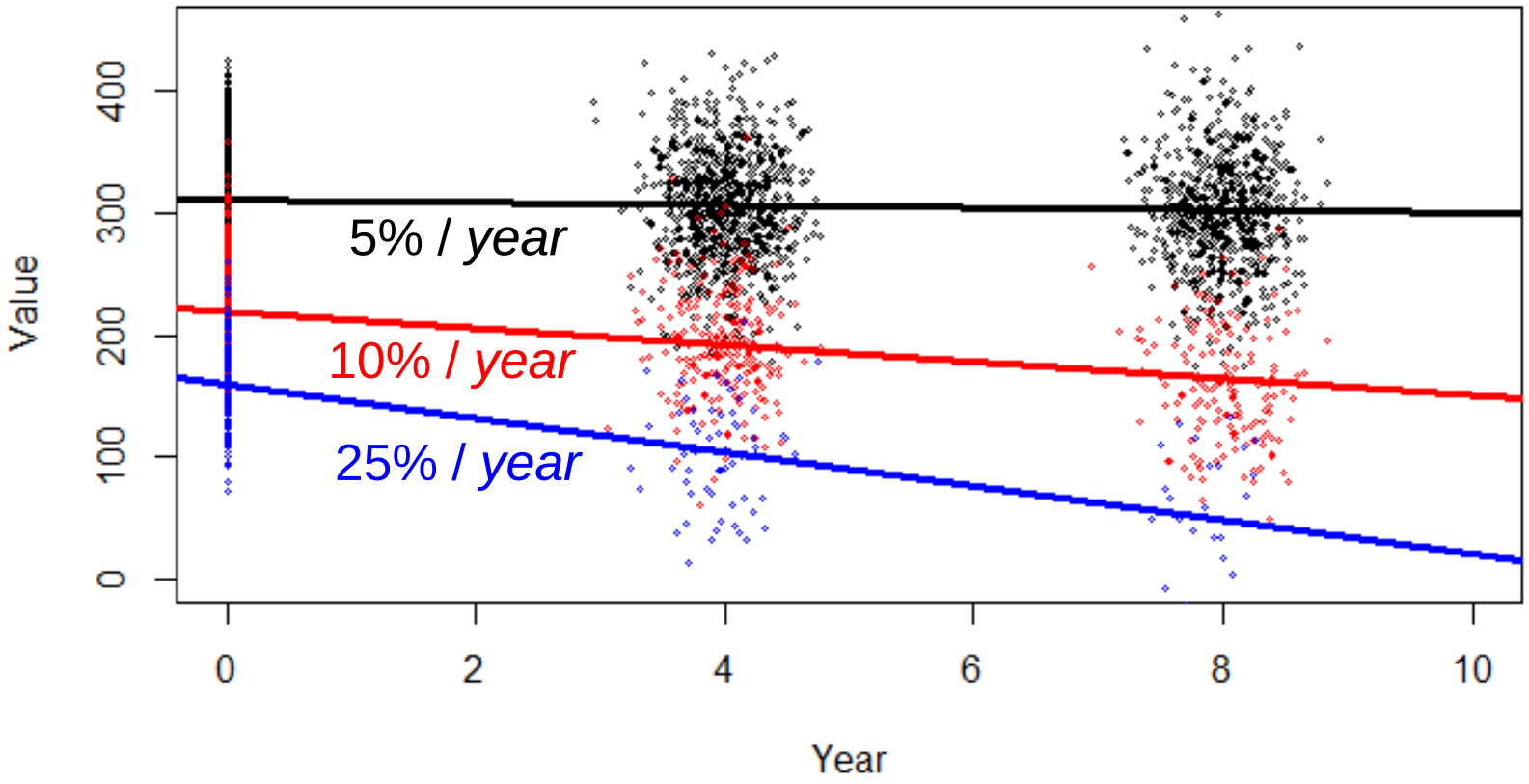
Individual trajectories within classes



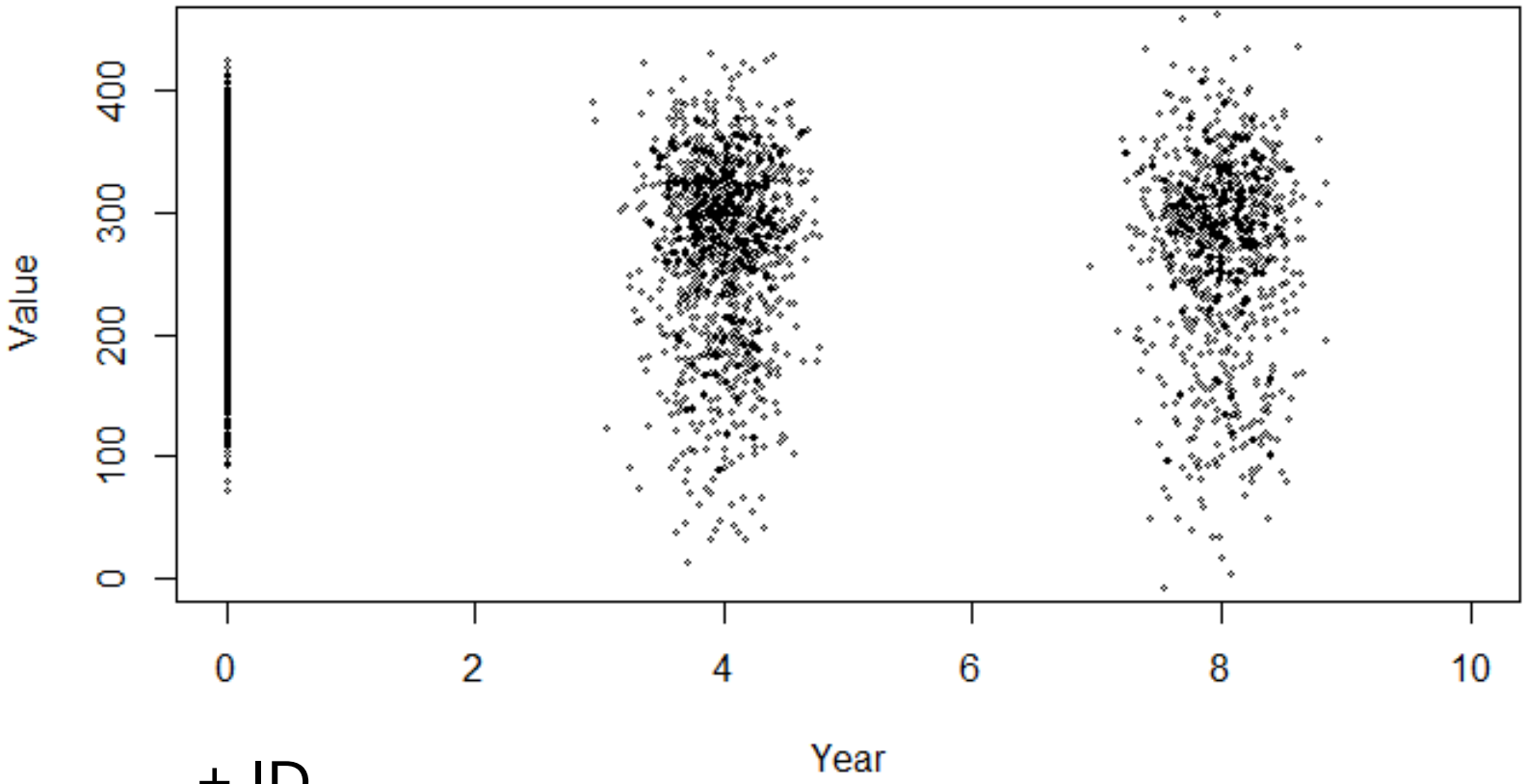
Measurements, including additional error



Inducing mortality



Available information for modeling



+ ID
+ event / censoring information

Expectations

Correct

- identification of trajectory classes
- estimation of mortality associated with each class
- estimation of class size/share/prevalence
- correct classification of individuals into classes

Model call

```
Jointlcmm(fixed= value~time, random= ~time, mixture= ~time, subject='id'
          , survival= Surv(censT,event) ~ 1, hazard="Weibull"
          , hazardtype="PH", ng=3, data=ds.alive)
```

id	time	value	censT	event
1	0.000000	366.7588	10.0000000	0
1	3.823100	341.4190	10.0000000	0
1	8.051839	294.5118	10.0000000	0
2	0.000000	358.8987	2.5436578	1
3	0.000000	331.8055	10.0000000	0
3	4.027998	281.7174	10.0000000	0
3	7.637723	344.7937	10.0000000	0

Model output

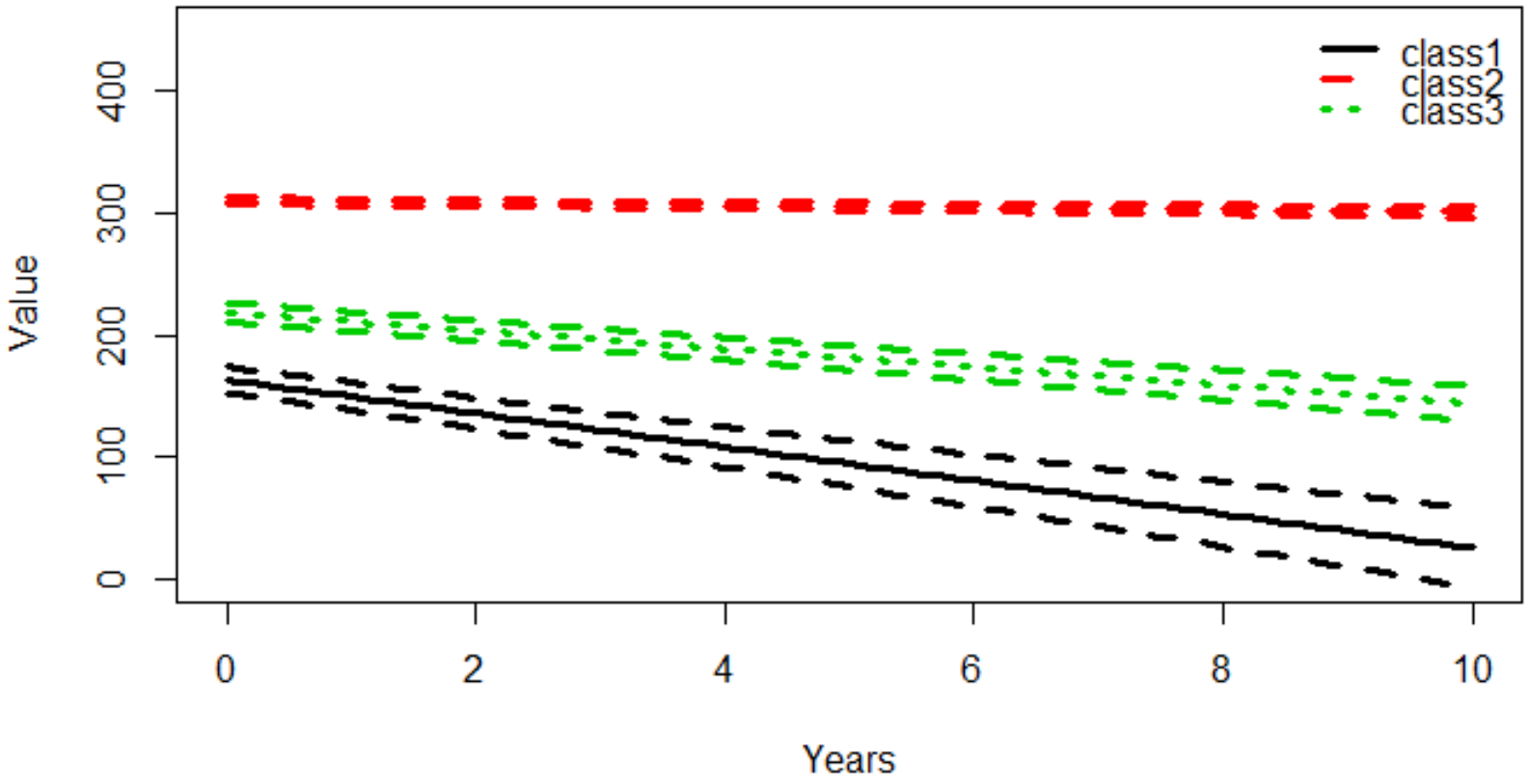
Parameters in the proportional hazard model:

	coef	Se	Wald	p-value
event1 +/-sqrt(Weibull1)	0.30483	0.01627	18.737	0.00000
event1 +/-sqrt(Weibull2)	1.03208	0.01780	57.968	0.00000
event1 SurvPH class1	1.06880	0.17238	6.200	0.00000
event1 SurvPH class2	-0.58125	0.13223	-4.396	0.00001

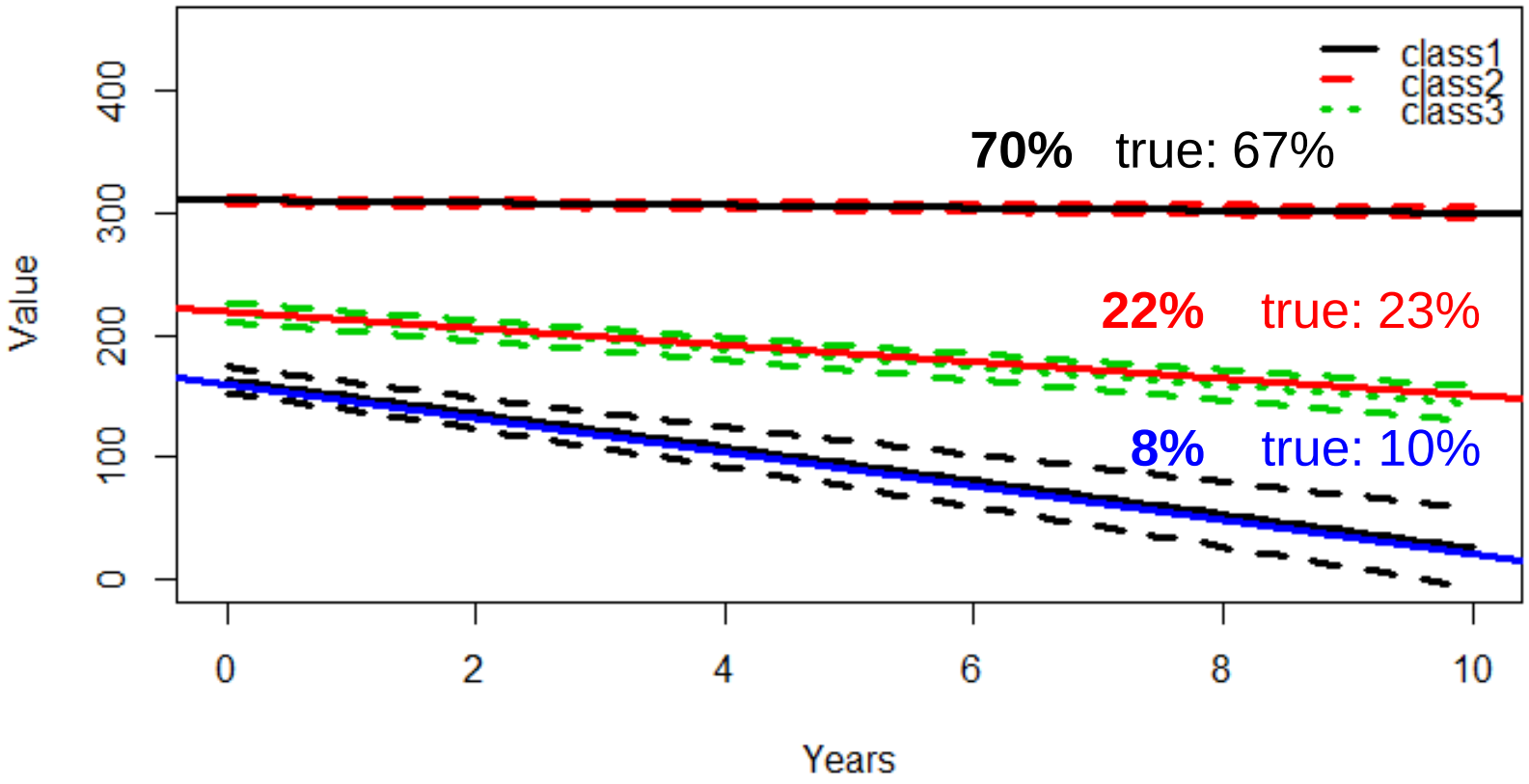
Fixed effects in the longitudinal model:

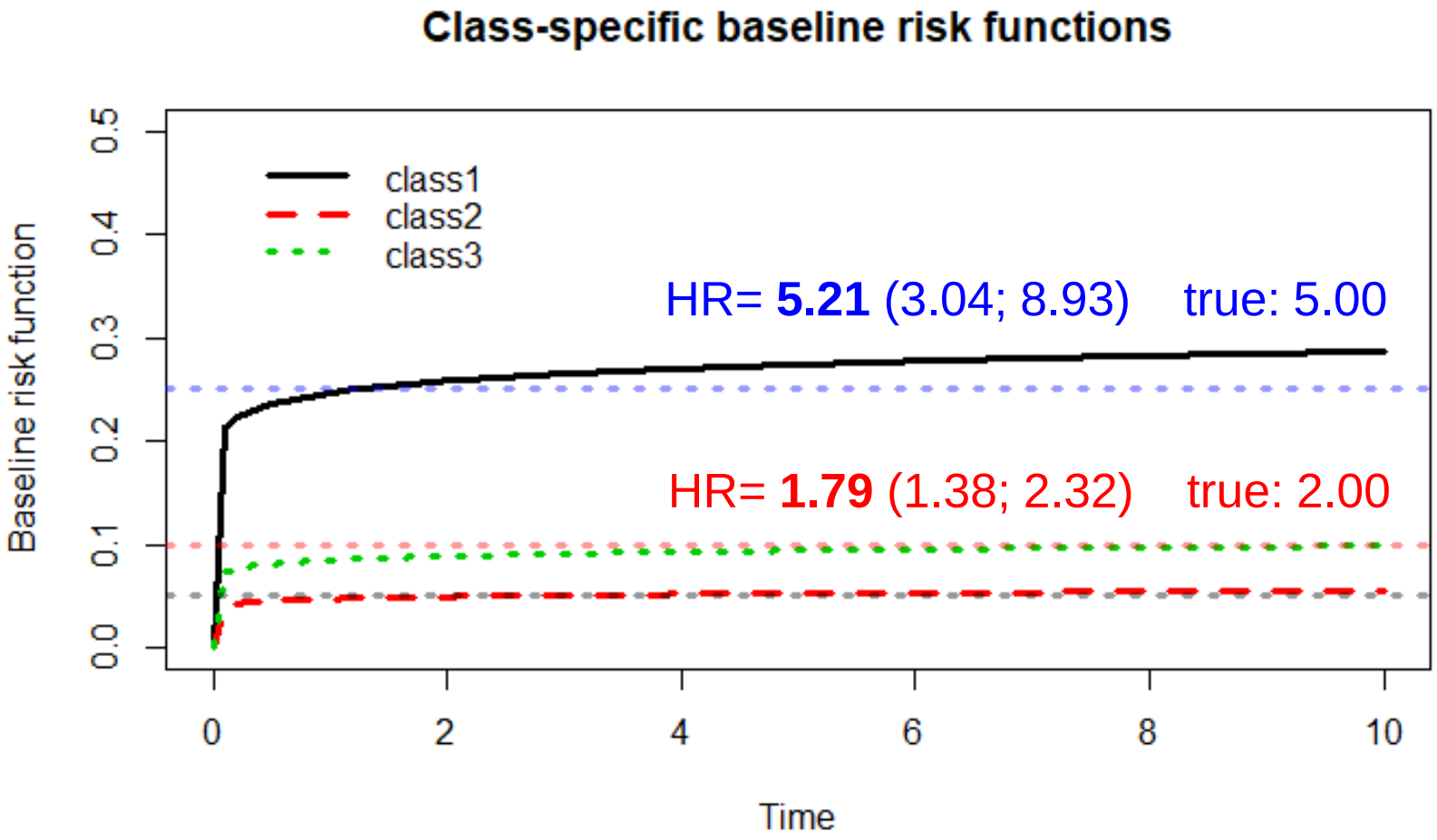
	coef	Se	Wald	p-value
intercept class1	163.41029	5.58704	29.248	0.00000
intercept class2	309.68170	1.47689	209.684	0.00000
intercept class3	218.41689	4.04109	54.049	0.00000
time class1	-13.75126	1.52353	-9.026	0.00000
time class2	-0.87654	0.22359	-3.920	0.00009
time class3	-7.42482	0.57591	-12.892	0.00000

Estimated trajectory classes



Estimated trajectory classes




Class-specific baseline risk functions
Baseline risk function
0.0
0.1
0.2
0.3
0.4
0.5
class1
class2
class3
HR= 5.21 (3.04; 8.93) true: 5.00
HR= 1.79 (1.38; 2.32) true: 2.00
0
2
4
6
8
10
Time

Classification performance

True class \ Estimated class	2	3	1
1	**977**	23	0
2	64	**243**	43
3	0	38	**112**

Agreement= 0.89
Kappa= 0.74

in those with baseline measurement only (n=370)

True class \ Estimated class	2	3	1
1	**162**	6	0
2	31	**50**	32
3	0	22	**67**

Agreement= 0.75
Kappa= 0.55

Expectations

Correct

- identification of trajectory classes
- estimation of mortality associated with each class
- estimation of class size/share/prevalence
- correct classification of individuals into classes

2. Depression trajectories and risk of cardiovascular events

KAROLA study

- Coronary heart disease patients participating in an in-patient rehabilitation program
- Followed after 1, 3, 6, 8, 10, 13, and 15 years
- Completed HADS at 1, 3, 6, 8 and 15 years
    - HADS Depression sub-scale 0-21 (7 items x 0-3) points (higher is worse)

Are depression trajectories associated with risk of subsequent CV-Events?

Study population

TABLE 1 Characteristics of the study population

	N	
Age (years), mean (*SD*)	1,109	59.4 (8.0)
Female sex, *n* (%)	1,109	171 (15.4)
Education <10 years, *n* (%)	1,109	663 (59.8)
Former/current smoker, *n* (%)	1,109	759 (68.4)
Body mass index, mean (*SD*)	1,109	26.9 (3.3)
Physical activity, *n* (%)		
Up to 3 hr per week	1,109	512 (46.2)
More than 3 hr per week	1,109	569 (51.3)
Not answered	1,109	28 (2.5)
History of myocardial infarction, *n* (%)	1,109	650 (58.6)
History of type 2 diabetes, *n* (%)	1,109	195 (17.6)
Clinical score (angiographic evaluation), *n* (%)		
0- or 1-vessel disease	1,109	287 (25.9)
2-vessel disease	1,109	291 (26.2)
3-vessel disease	1,109	476 (42.9)
Unknown	1,109	55 (5.0)
Depression score, mean (*SD*)		
Baseline	1,093	4.4 (3.5)
1-year follow-up	975	4.7 (3.8)
3-year follow-up	809	4.7 (3.7)
6-year follow-up	784	4.3 (3.7)
8-year follow-up	711	4.4 (3.7)
15-year follow-up	408	4.6 (3.6)
Follow-up for CV events/non-CV mortality		
Follow-up time (years), median (Q1; Q3)[a]	1,109	14.8 (13.3; 14.9)
CV events, n (%)	1,109	324 (29.2)
Non-CV mortality, n (%)	1,109	147 (13.3)
Lost before last follow-up, n (%)	1,109	232 (20.9)

Peter et al., 2020

Selecting the number of classes

```
> summarytable(lcmm1.dep, lcmm2.dep, lcmm3.dep, lcmm4.dep, lcmm5.dep)
           G    loglik npm      BIC    %class1  %class2  %class3  %class4  %class5
lcmm1.dep  1 -14102.37  34 28443.12 100.000000                                     
lcmm2.dep  2 -14021.68  39 28316.80  87.466186 12.53381                           
lcmm3.dep  3 -13971.88  44 28252.26   4.598738 84.49053 10.91073                  
lcmm4.dep  4 -13938.72  49 28220.98   3.246168 69.61226 23.80523 3.336339        
lcmm5.dep  5 -13948.64  54 28275.88   4.508566  0.00000 82.14608 5.410280 7.935077
```

Model call

```
Jointlcmm(fixed= dep~ time, random= ~time, mixture= ~time, subject='ID',
          survival = Surv(tentry,texit,event)~ cause(baseage) + cause(SEX) +
          cause(schooledu) + cause(smk) + cause(bmie) + 
          cause(KAHWOCH.1) + cause(KAHWOCH.2) + cause(HERZINF) + cause(dm) +
          cause(GEFAESSN.2) + cause(GEFAESSN.3) + cause(GEFAESSN.4),
          hazard="Weibull", hazardtype=c("PH", "PH"), ng=4, data=ds)
```

Results

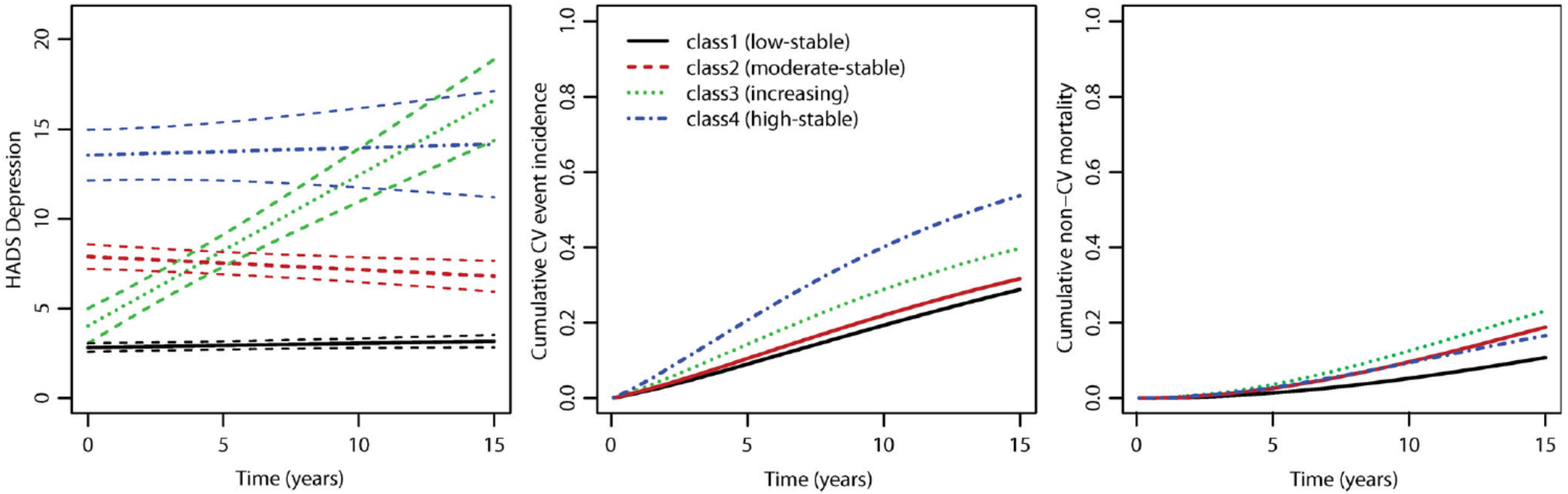


FIGURE 1 Identified anxiety and depression trajectories and event incidence. Left panels: identified anxiety and depression trajectory classes (means with corresponding 95% confidence bands); middle panels: cumulative CV event incidence by follow-up time for mean covariate levels; right panels: cumulative non-CV mortality for mean covariate levels. CV, cardiovascular; HADS, Hospital Anxiety and Depression Scale

Results

TABLE 2 Risk of subsequent CV events and non-CV mortality by trajectory class (n = 1,109)

		CV events		Non-CV mortality	
Trajectory	Class share in %	Cumulative 15-year incidence[a]	HR (95% CI)[a]	Cumulative 15-year mortality[a]	HR (95% CI)[a]
Depression					
Class 1 (low-stable)	69.6	28.8%	(Ref.) 1.00	10.7%	(Ref.) 1.00
Class 2 (moderate-stable)	23.8	31.7%	1.18 (0.34; 4.07)	18.8%	1.94 (0.21; 17.5)
Class 3 (increasing)	3.3	39.7%	1.65 (0.44; 6.15)	23.1%	2.76 (0.31; 24.7)
Class 4 (high-stable)	3.3	53.7%	2.47 (1.35; 4.54)	16.6%	2.26 (0.77; 6.65)

Abbreviations: CI, confidence interval; CV, cardiovascular; HR, hazard ratio.

[a]Adjusted for baseline sex, age, education, smoking status, body mass index, physical activity, history of myocardial infarction, history of type 2 diabetes, and angiographic evaluation.

3. Physical activity trajectories and mortality

ActiFE Ulm study

- population-based cohort study in older people ≥65 years, randomly selected in the Ulm region
- activPAL was worn for seven consecutive days at baseline and follow-up
  - Baseline 2009/10: 1335
  - FU 1 2012/13: 790
  - FU 2 2017/18: 508

    -> 16 761 measurement days in total

Are objectively measured physical activity trajectories associated with mortality?

Important: need to correct for weather conditions (Klenk et al. 2012) and Sunday-effect (Klenk et al., 2019)

Baseline characteristics

Table 1: Baseline characteristics for the total cohort and stratified by trajectory class

	Total (n=1,406)
Age (years), mean (SD)	75.4 (6.52)
Female, n (%)	616 (43.8)
Duration of school education ≤9 yrs, n (%)	788 (56.1)
Former/current smoker, n (%)	671 (49.5)
Daily alcohol consumption, n (%)	427 (30.4)
BMI (kg/m²), mean (SD)	27.6 (4.16)
Habitual gait seed (m/s), mean (SD)	1.04 (0.33)
Hand grip strength (kg), mean (SD)	32.2 (11.1)
Walking duration (min), mean (SD)	104.1 (40.3)
Number of deaths, n (%)	390 (27.7)
Follow-up duration (years), median (Q1-Q3)	9.10 (8.55-9.46)

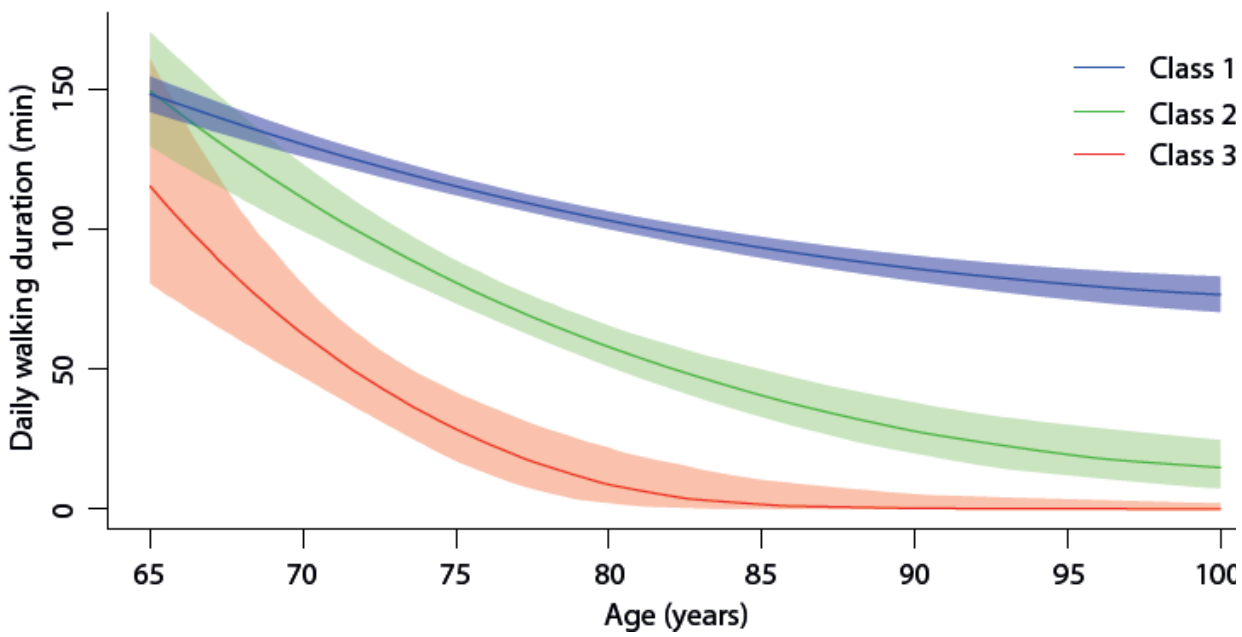
Data

	sid	walkd	m_age	sday	temp	sex	tentry	texit	death2019
1	20034	96.995000	70.13973	0	7.95	1	70.13973	79.35890	0
2	20034	139.021667	70.14247	0	5.36	1	70.13973	79.35890	0
3	20034	98.745000	70.14521	1	6.04	1	70.13973	79.35890	0
4	20034	106.675000	70.14795	0	8.86	1	70.13973	79.35890	0
5	20034	153.481667	70.15068	0	10.61	1	70.13973	79.35890	0
6	20034	138.410000	70.15342	0	7.38	1	70.13973	79.35890	0
7	20034	82.668333	73.40822	0	-1.54	1	70.13973	79.35890	0
8	20034	111.035000	73.41096	0	-1.60	1	70.13973	79.35890	0
9	20034	87.820000	73.41370	0	-4.05	1	70.13973	79.35890	0
10	20034	58.900000	73.41644	0	-5.76	1	70.13973	79.35890	0
11	20034	66.213333	73.41918	0	-5.06	1	70.13973	79.35890	0
12	20034	75.576667	73.42192	1	-4.07	1	70.13973	79.35890	0
13	20034	90.933333	73.42466	0	-2.79	1	70.13973	79.35890	0
14	20034	65.501667	77.62740	0	8.55	1	70.13973	79.35890	0
15	20034	63.966667	77.63014	0	12.26	1	70.13973	79.35890	0
16	20034	51.816667	77.63288	0	14.38	1	70.13973	79.35890	0
17	20034	75.048333	77.63562	0	14.43	1	70.13973	79.35890	0
18	20034	56.315000	77.63836	1	13.76	1	70.13973	79.35890	0
19	20034	71.115000	77.64110	0	13.40	1	70.13973	79.35890	0
20	20034	131.445000	77.64384	0	14.11	1	70.13973	79.35890	0

Model call

```
Jointlcmm(fixed= walkd~ m_age + sday + temp,random=~ m_age, mixture=~ m_age,
          subject='sid', link="3-quant-splines",
          survival = Surv(tentry,texit,death2019)~ sex,
          hazard="Weibull", hazardtype="PH",ng=3,data=ds)
```

Results



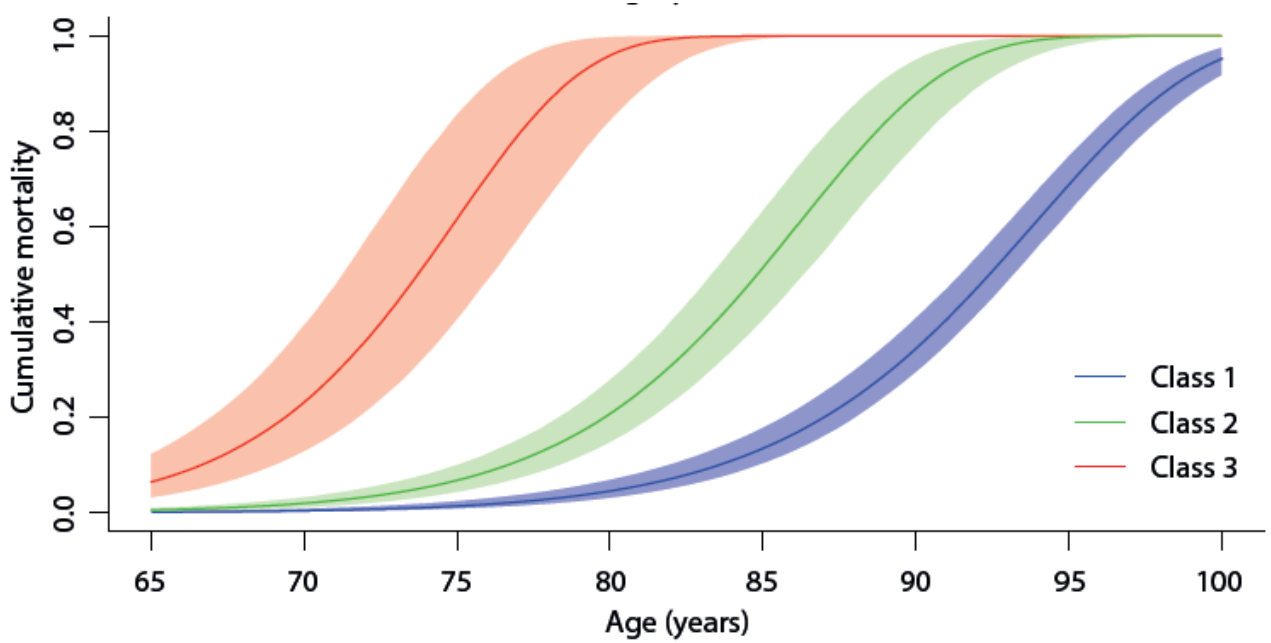


Figure 1: Identified walking duration trajectories over age (left panel) and corresponding cumulative mortality over age (right panel). Trajectories are adjusted for measurement day (Sunday vs. other weekdays) and maximum daily temperature and presented for a non-Sunday weekday of average maximum temperature (7.8 °C).
Mortality curves are adjusted for sex and presented for a population of 50% men and 50% women.

Results

Table 2: Mortality by trajectory class (N=1,406, N_{events}=390)

	Class share in % [1]	Life expectancy, median age in years (95%-CI) [2]	Hazard ratio (95%-CI) [3]
Class 1	87.1	92.4 (91.4; 93.4)	1.0 (ref)
Class 2	11.2	84.8 (83.4; 86.4)	4.96 (3.37; 7.32)
Class 3	1.7	73.8 (71.1; 76.2)	68.7 (29.0; 162.6)

[1] posterior classification based on longitudinal and time-to-event data

[2] for a population of 50% men and 50% women

[3] adjusted for age and sex

Associations with class membership

Table 3: Relative probability and 95% confidence intervals of class membership by baseline characteristics (n=1,398) and biomarkers (n=1,345), adjusted for age and sex

		RR (95%-CI)	
	Class 1	Class 2	Class 3
Clinical characteristics			
Education < 10 years	1.00 (Ref.)	1.05 (0.83; 1.32)	0.84 (0.36; 1.94)
Former/current smoker	1.00 (Ref.)	1.27 (0.85; 1.90)	2.64 (0.98; 7.13)
Daily alcohol consumption	1.00 (Ref.)	1.18 (0.92; 1.53)	1.41 (0.58; 3.43)
Body mass index (per 2.5 kg/m²)	1.00 (Ref.)	1.17 (1.11; 1.23)	1.16 (0.95; 1.42)
Biomarkers (per standard deviation)			
CRP [mg/L]	1.00 (Ref.)	1.23 (1.10; 1.37)	2.06 (1.42; 3.01)
IL-6 [pg/ml]	1.00 (Ref.)	1.27 (1.14; 1.40)	2.26 (1.65; 3.10)
NT pro BNP [pg/ml]	1.00 (Ref.)	1.31 (1.16; 1.47)	2.34 (1.72; 3.17)
hs Troponin T [ng/L]	1.00 (Ref.)	1.35 (1.21; 1.51)	2.36 (1.77; 3.15)
hs Troponin I [ng/L]	1.00 (Ref.)	1.23 (1.12; 1.35)	1.81 (1.46; 2.25)
Cholesterol [mmol/L]	1.00 (Ref.)	0.83 (0.73; 0.94)	0.58 (0.37; 0.92)
HDL cholesterol [mmol/L]	1.00 (Ref.)	0.80 (0.70; 0.91)	0.43 (0.24; 0.76)
LDL cholesterol [mmol/L]	1.00 (Ref.)	0.85 (0.76; 0.96)	0.62 (0.39; 0.98)
γ-GT [U/l]	1.00 (Ref.)	1.23 (1.10; 1.38)	2.08 (1.39; 3.10)
Glucose [mg/dl]	1.00 (Ref.)	1.13 (1.02; 1.25)	1.42 (1.06; 1.91)
Creatinine serum [µmol/L]	1.00 (Ref.)	1.20 (1.10; 1.30)	1.48 (1.23; 1.77)
Cystatin C serum [mg/L]	1.00 (Ref.)	1.22 (1.11; 1.33)	1.52 (1.29; 1.79)
GDF-15 [pg/ml]	1.00 (Ref.)	1.36 (1.22; 1.51)	2.27 (1.69; 3.06)

4. Pros & Cons

Pro

- Use of all information
  - also including individuals with one measurement only
  - including all events

  leading to
  - less selection bias
  - higher power

- Adjustment in both sub-models possible

Contra

- Computationally expensive

- Small classes are no exception

- **Adjustment sets are fixed**