GÖTEBORGS UNIVERSITET

Statistical methods in life-course epidemiology

Monica Leu Agelii, EPSO

Currently happening in research

- Increasing emphasis on the fetal and childhood triggers of chronic complex disease
- Increasing awareness about how exposures across the life-course interact to influence later health
- Large-scale studies with a range of structural, social and cultural ***determinants*** of health-related outcomes
  - Often the data originates from clusters (schools, multi-site interventions) or nested clusters (hierarchies: schools within countries)

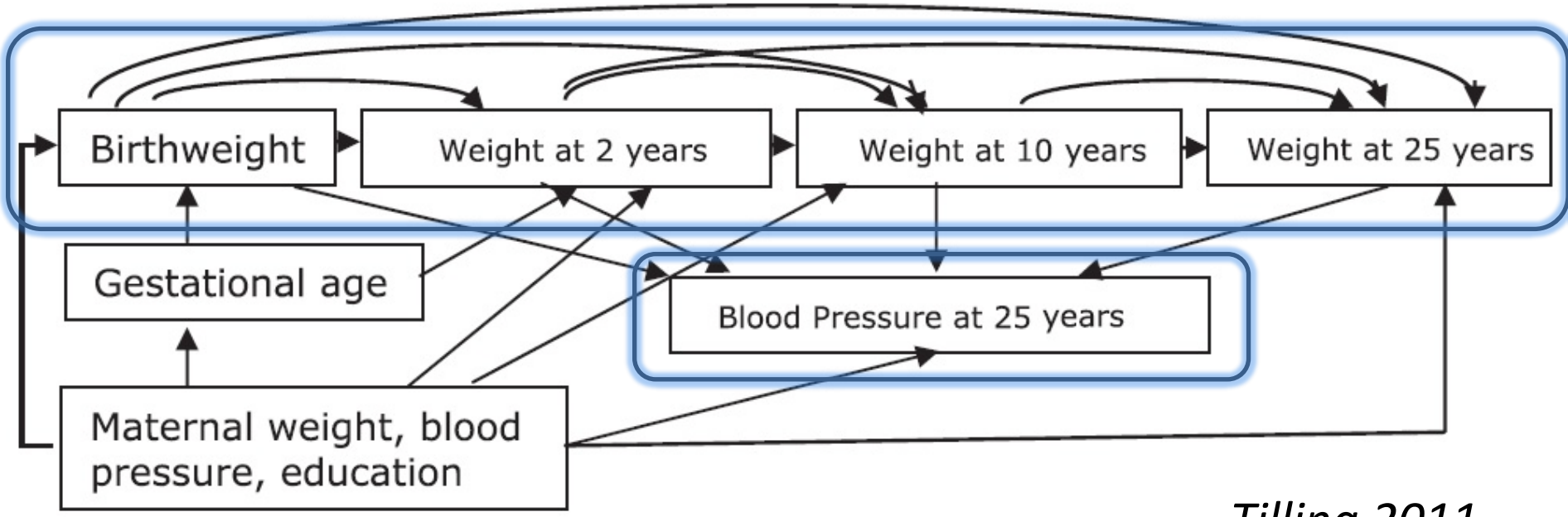
Current needs in health research

- Statistical methods that can deal with data dependencies
  - Repeated measures (longitudinal studies): change over time in both **continuous** and **dichotomous** outcomes
  - Data **clustering** (observations in the same cluster are more similar than observations in separate clusters)
  - Repeated measures **+** clustering

Current needs in health research

- Dependencies between repeated observations
- Clustering
- Measurement error varying over time
- Unequal number of measurements between individuals
- Varying times of measurement (ex. different ages)

Life course example: weight (changes) and blood pressure



Tilling 2011

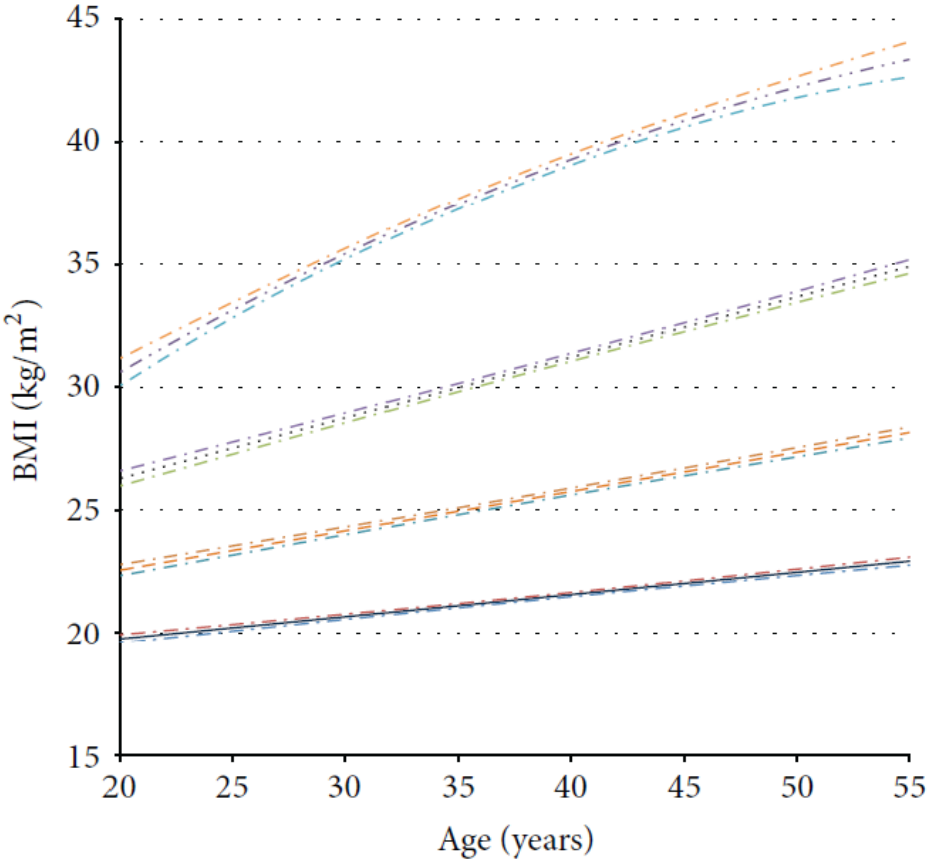
- Can study the development of weight trajectories
- Or study the relationship between weight (as a time-dependent covariate) and blood pressure at age 25 yrs

Modelling strategies

The model depends on the *a-priori* hypothesis

- Regression with covariates
- Individual growth modelling (changes in a characteristic over time); trajectories
- Hazard modelling (Cox regression)
- Structural equation modelling (SEM)
- Propensity score analysis (PSA)
- Others…

Individual growth modelling



Growth: not only about *developmental growth (ex: height in children)* but rather modelling the evolution of characteristic (ex: blood pressure in adulthood)

Individual growth models

- Multivariate repeated measures (only with fixed effects, or with fixed & random effects): MIXED models
  - Accommodate for clusters through the random effects
- Latent class analysis
- Latent growth-curve modelling
- Growth mixture modelling

MIXED models

For a continuous outcome:

- Provide estimates of average level in the outcome at a given time (intercept)
- Estimate the average linear rate of change over time (slope)
- Estimate the average non-linear rate of change (quadratic, cubic)

MIXED models

- **Advantages**
  - Use all available data, without restricting only to individuals with complete data
  - Data can be measured at varying times or ages
  - Allow for individual differences in the intercept and slope (random intercept, random slope models – also referred to as *variance components*)

MIXED models

- **Limitation:** Models assume that all individuals in the population have the same pattern (direction) of change
  - **OK** in many situations: ex. modelling height during childhood (height expected ↗ in all children)
  - **Not always OK**: ex. part of the population BMI↗ over study period and part ↘ => the average pattern shows a constant BMI
    - When in this situation go with latent class growth models

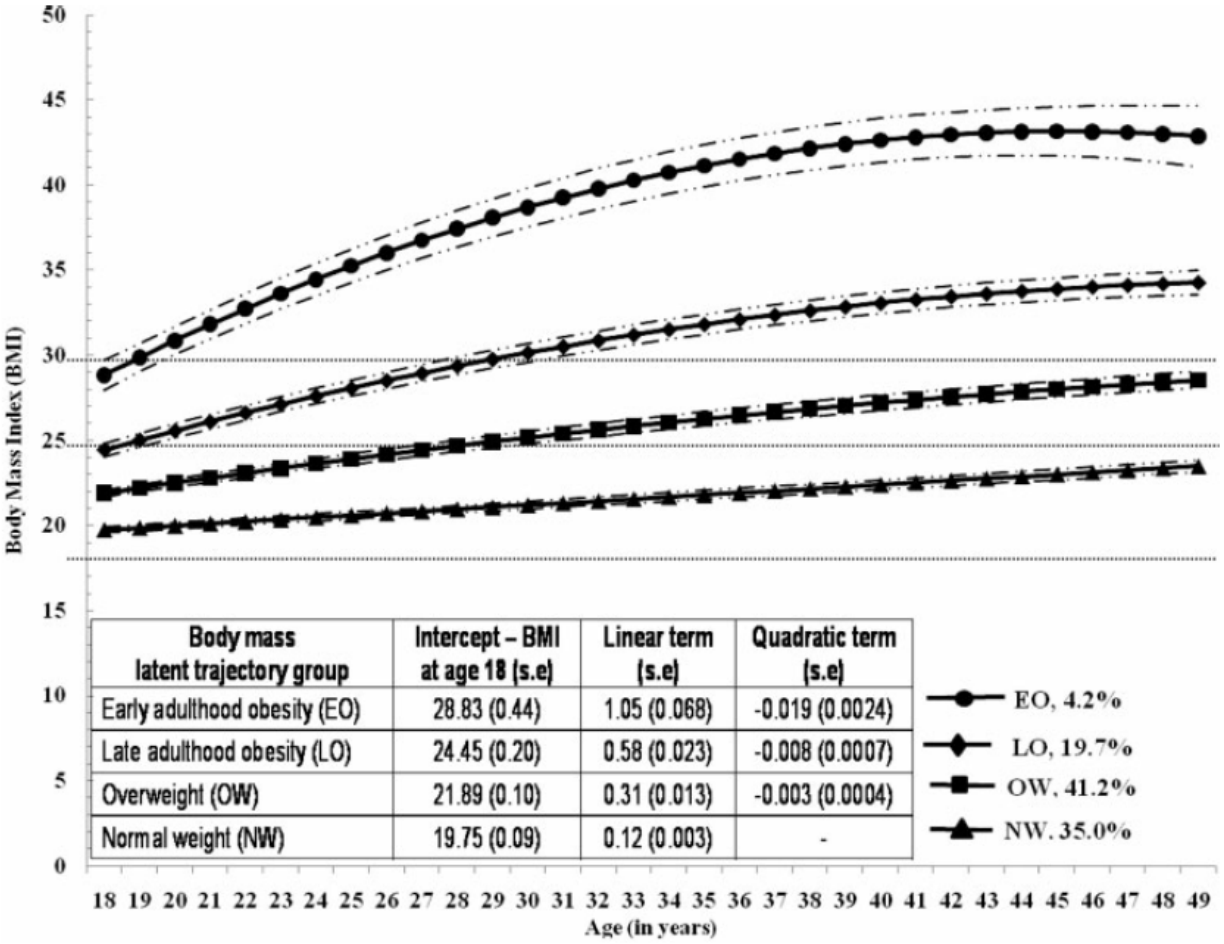
Latent class growth models (LCGM)

- Allow subpopulations (classes) to have their own average pattern of growth - trajectories
- Each subject is given a probability of belonging to each class, p_{class} and is placed in the class where s/he has the highest p_{class}
- Detects covariates associated with class membership
- Further on: we can examine associations between class membership and later outcome

LCGM concepts

- Have to specify the number of trajectories
  - Select the model with number of trajectories that best fit the data (BIC)
- Trajectories can be of various shapes (linear, cubic..)
- Each trajectory should hold at least ~5% of subjects
- Ideally, at least 300-500 subjects in the data
  - Smaller sample size limits the number of identifiable trajectories

LCGM Example: Several BMI trajectories from 18 to 49 years, Østbye 2011



Body mass latent trajectory group	Intercept – BMI at age 18 (s.e)	Linear term (s.e)	Quadratic term (s.e)
Early adulthood obesity (EO)	28.83 (0.44)	1.05 (0.068)	-0.019 (0.0024)
Late adulthood obesity (LO)	24.45 (0.20)	0.58 (0.023)	-0.008 (0.0007)
Overweight (OW)	21.89 (0.10)	0.31 (0.013)	-0.003 (0.0004)
Normal weight (NW)	19.75 (0.09)	0.12 (0.003)	-

CVD: Yes/No

CVD: Yes/No

CVD: Yes/No

CVD: Yes/No

My LCGM example

BMI trajectories from birth until study enrollment as predictors of vitamin D deficiency

Working tool: Proc TRAJ

BMI trajectories <-> Vitamin D deficiency

The ongoing conundrum: Higher BMI leads to vitamin D deficiency? or vitamin D deficiency leads to higher BMI?

Working hypothesis: children on higher BMI trajectories are at increased risk of vitamin D deficiency

Background on the data

- Perinatal data from 7 countries
- BMI measurements from birth until ~ 5-6 years of age
  - 1000 children with 25(OH)D at enrollment
- Had to be harmonized regarding frequency and order of the measurements
  - Initially between 1-40 measurements/ subject, some of these very close in time
  - Order: 1st measurement for a subject (say at 24 mo's) could be at the time of the 5th measurement for other subjects
- Had to be prepared (wide format) for Proc TRAJ

Harmonizing the data

- Age (months) was first categorized into classes

  if (age_time <=**6**) then agec=**1**;
  if (**6**< age_time <=**12**) then agec=**2**;
  if (**12**< age_time <=**18**) then agec=**3**;
  …..

- From each age class, one measurement was randomly selected
- Resulted in maximum 13 repeated measurements
- Each subject kept one BMI measurement per age-class (or recorded as missing) and the actual age at measurement

The data – as Proc TRAJ likes it

VIEWTABLE: Work.Trans_birthcard

	ID-Number	age_time1	BMI1	age_time2	BMI2	age_time3	BMI3	age_time4	BMI4
118	1610593	6	16.05	12	16.19	18	15.46	.	.
119	1610594	6	18.83	12	18.13	18	16.56	.	.
120	1610598	5	15.92	12	17.65	.	.	19	16.44
121	1610601	6	16.84	12	15.46	18	17.04	.	.
122	1610608	6	15.36	12	15.83	18	15.30	.	.
123	1610609	.	.	12	18.31	.	.	20	19.14
124	1610610	5	15.94	11	17.65	18	16.24	.	.
125	1610611	.	.	11	19.31	.	.	.	.
126	1610627	6	17.83	.	.	13	.	.	.
127	1610632	6	14.61	.	.	13	15.12	.	.
128	1610633	6	16.50	.	.	13	15.22	.	.
129	1610634	6	15.46	12	15.38	.	.	19	15.56
130	1610636	6	14.25	12	15.35	18	14.69	.	.
131	1610639	5	17.26	12	16.23	18	15.90	.	.
132	1610642	.	.	12	15.69	.	.	.	.
133	1610643	6	19.54	12	19.59	18	17.98	.	.
134	1610648	5	18.27	.	.	13	15.19	20	14.23
135	1610650	6	16.98	12	16.18	.	.	19	16.17
136	1610657	6	17.78	12	15.25	18	15.50	.	.
137	1610658	6	14.95	10	.	.	.	.	.
138	1610664	.	.	12	17.78	17	18.34	.	.
139	1610667	6	15.79	12	14.68	.	.	.	.
140	1610671	6	15.83	12	15.53	18	14.94	.	.
141	1610681	6	18.13	12	17.94	18	17.56	.	.
142	1610682	6	17.58	12	17.77	.	.	19	16.95
143	1610685	6	18.12	10	17.42	.	.	20	16.45
144	1610690	6	17.49	12	16.00	18	17.88	.	.
145	1610705	.	.	7	19.00	18	17.64	.	.
146	1610713	6	16.06	12	15.05	18	14.70	.	.

PROC TRAJ

- **Not actually a regular SAS procedure, it's a macro.**
- **Freely available for download, with instructions for where to place it**
  https://www.andrew.cmu.edu/user/bjones/download.htm

```
proc traj data=birthcard OUT=OF OUTPLOT=OP
        OUTSTAT=OS OUTEST=OE
        ITDETAIL ALTSTART CI95M;
   id ID_no;
   var bmi1-bmi5 bmi8-bmi10;
   indep age_time1- age_time5 age_time8-age_time10;
   model CNORM; max 30; ngroups 4; order 3 3 3 3;
run;
```

How many trajectories and what kind of shape?

- Start with 1 trajectory, quadratic => all β's sign.
- … cubic => better fit (BIC) and all β's sign.
- 2 trajectories, cubic => better fit
- ….
- 4 trajectories, cubic => better fit
- 5 trajectories => no improvement (5th trajectory did not exist)

Output files:
Trajectory membership

Assigned trajectory

Probability to belong to each trajectory

VIEWTABLE: Work.Of

	ID-Number	BMI1	BMI2	BMI3	age_time1	age_time2	age_time3	Group 1 Probability	Group 2 Probability	Group 3 Probability	Group 4 Probability	Group
1	1110065	10.89	.	.	0	.	.	0.764954	0.005827	0.000000	0.229219	1
2	1110070	15.29	16.06	.	6	12	.	0.310460	0.678210	0.001263	0.010067	2
3	1110095	14.96	14.63	15.29	2	11	15	0.729322	0.270647	0.000003	0.000028	1
4	1130014	16.39	16.61	.	6	12	.	0.057709	0.901506	0.017105	0.023681	2
5	1130022	11.33	16.83	.	0	11	.	0.673130	0.029546	0.000000	0.297324	1
6	1130025	13.2	17.27	.	1	12	.	0.407536	0.408591	0.000172	0.183701	2
7	1130085	19.03	18.37	.	6	12	.	0.000036	0.100486	0.882375	0.017102	3
8	1130090	19.03	19.58	.	6	12	.	0.000002	0.023051	0.961489	0.015457	3
9	1130110	.	16.83	.	.	12	.	0.147844	0.752226	0.074375	0.025556	2
10	1130135	17.93	.	.	5	.	.	0.007430	0.556718	0.414432	0.021421	2
11	1130140	15.62	.	.	1	.	.	0.081706	0.877154	0.030496	0.010644	2
12	1130158	16.83	16.17	.	5	11	.	0.037754	0.925343	0.020859	0.016043	2
13	1130159	14.74	15.4	.	1	12	.	0.383125	0.613744	0.000293	0.002838	2
14	1130169	15.84	.	.	0	.	.	0.040209	0.904577	0.053211	0.002003	2
15	1130174	.	17.38	.	0	11	.	0.071071	0.724402	0.150015	0.054512	2
16	1130179	17.49	16.83	.	6	12	.	0.010728	0.854033	0.107924	0.027315	2
17	1130187	17.71	16.94	.	5	11	.	0.004433	0.789973	0.183452	0.022142	2
18	1130205	.	.	16.5	.	.	16	0.175805	0.798876	0.024842	0.000477	2

Output files:
Trajectory estimates

Beta coefficients for each of the 4 trajectories

REMOVED

Estimated mean and 95% CI for each trajectory and by time point (8 points here)

REMOVED

Useful literature

- **Andruff et al, 2009. Latent class growth modelling: a tutorial.** (PROC TRAJ tutorial)
- Bub and Ferretti, 2014. Cutting-edge statistical methods for a life-course approach.
- Hardy et al, 2003. Birthweight, childhood social class, and change in adult blood pressure in the 1946 British birth cohort.
- Liu et al, 2013. Social position and chronic conditions across the lifespan and risk of stroke: a life-course epidemiological analysis of 22847 American adults in ages over 50.
- Tilling et al, 2011. Commentary: Methods for analyzing life course influences on health – untangling complex exposures
- Østbye et al, 2011. Body mass trajectories in adulthood: results from the National Longitudinal Survey of Youth 1979 Cohort (1981-2006)