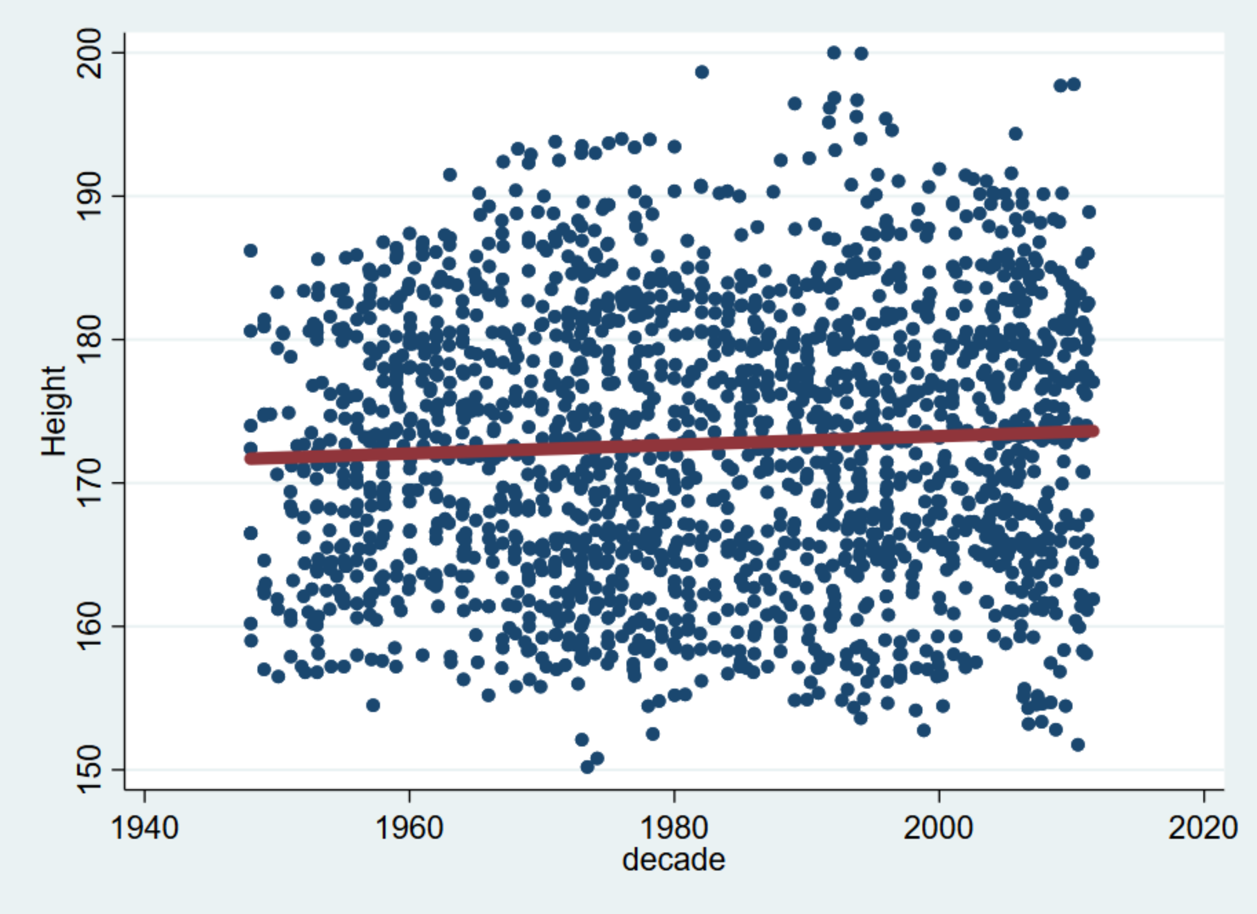
A beginners guide to (growth trajectory analysis using) **multilevel models**

Will Johnson

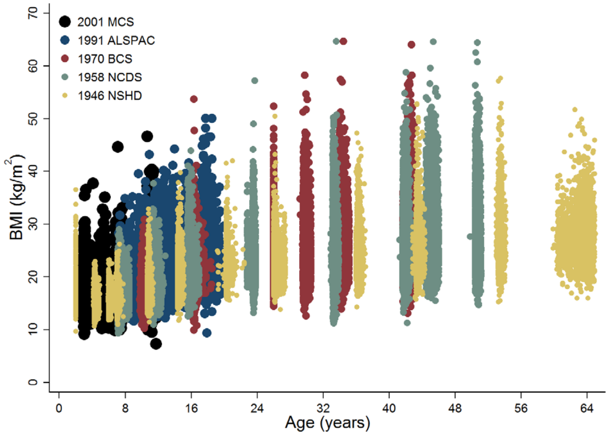
W.O.Johnson@lboro.ac.uk

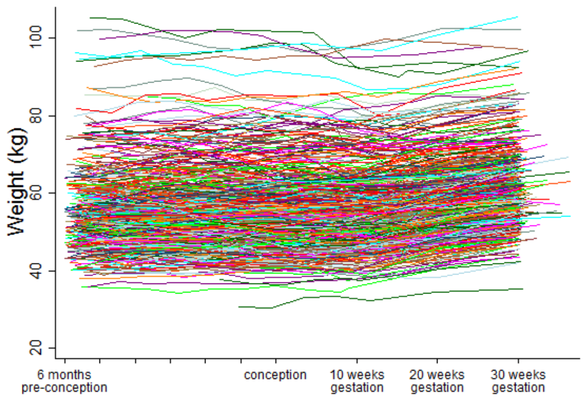
School of Sport, Exercise and Health Sciences, Loughborough University

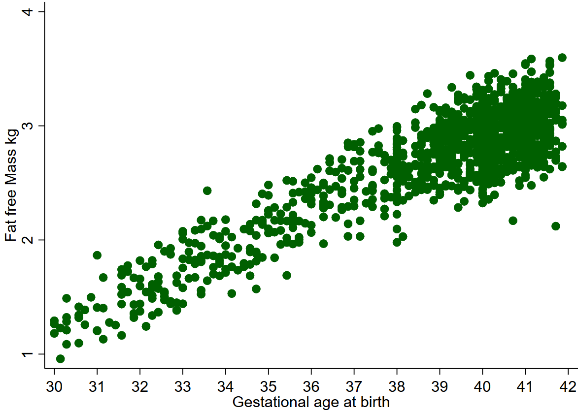
What is a trajectory?

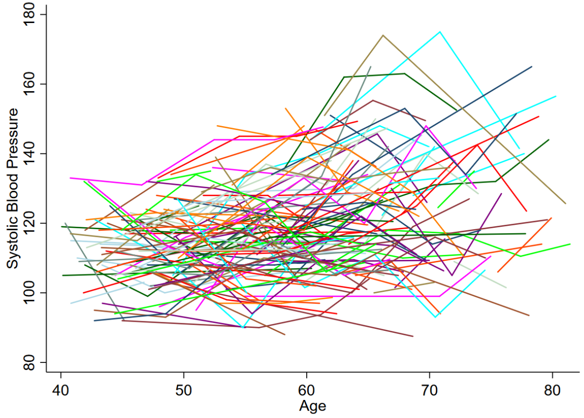


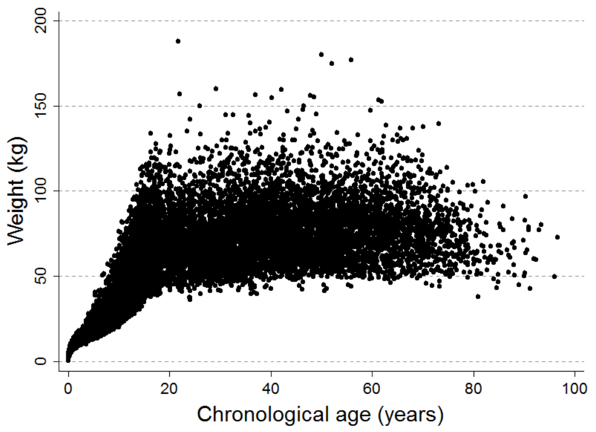
Longitudinal data

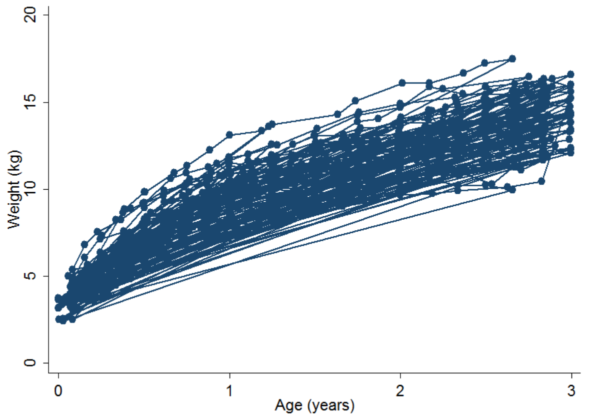








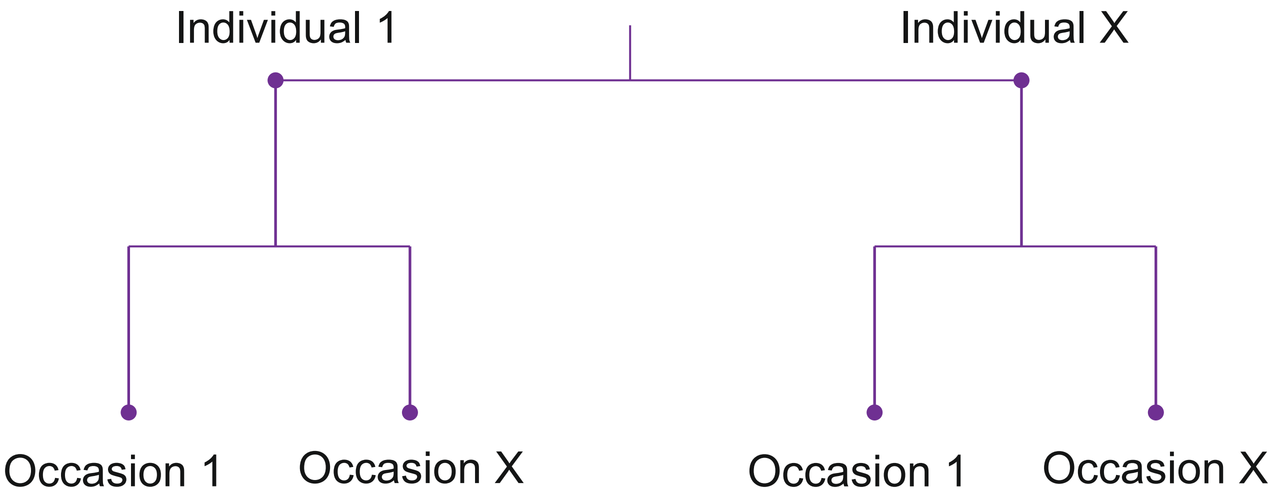




Structure of longitudinal data

clustered within
individuals (level 2)

serial measurement
occasions (level 1)



Structure of longitudinal data

	Occasion			
Individual	1	2	3	4
1	y_{11}	y_{21}		y_{41}
2	y_{12}	y_{22}	y_{32}	y_{42}
3	y_{13}		y_{33}	y_{43}

Individual	Occasion	
1	1	y_{11}
1	2	y_{21}
1	3	
1	4	y_{41}
2	1	y_{12}
2	2	Y_{22}
2	2	Y_{32}
2	2	Y_{42}
3	1	Y_{13}
3	2	
3	2	Y_{33}
3	2	Y_{43}

$\boldsymbol{y_{ij}}$ is the response at occasion $\boldsymbol{i}$ for individual $\boldsymbol{j}$

Other examples of multilevel data

How many upper-level units?

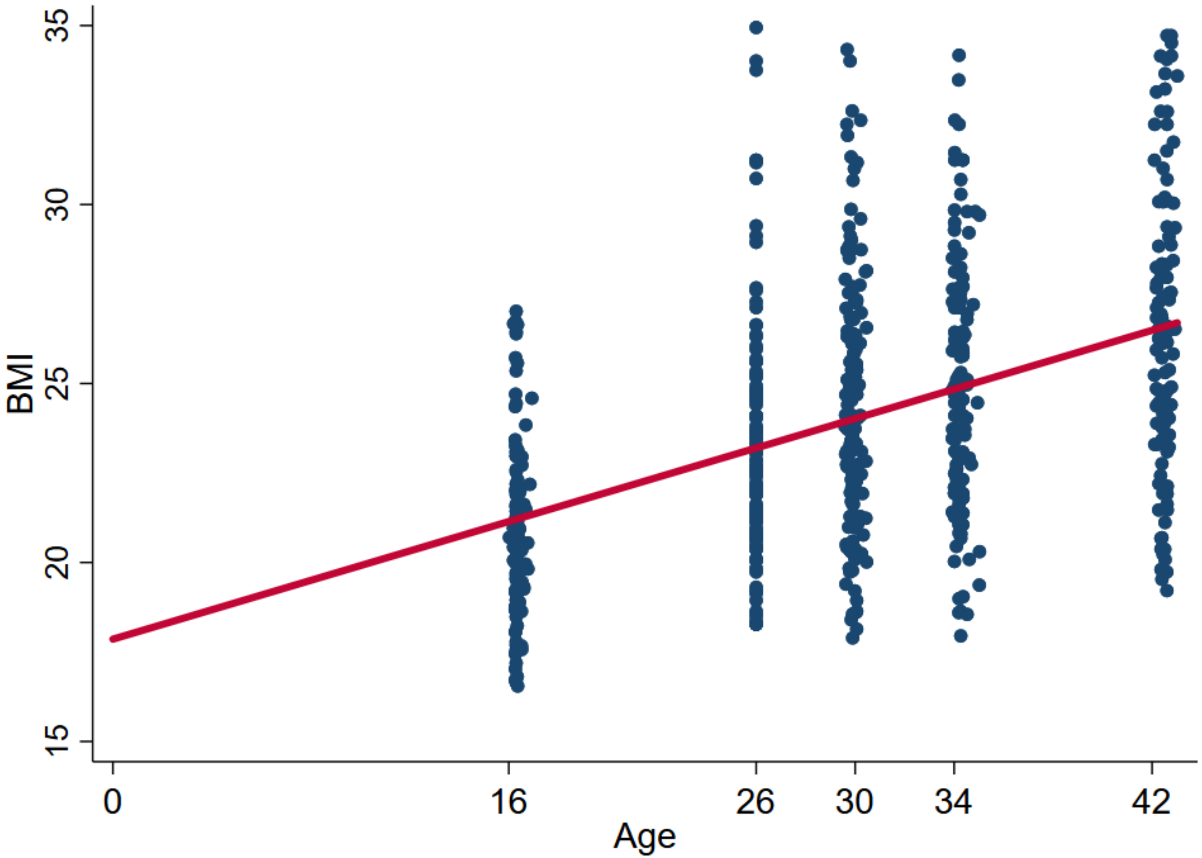
Elff, Martin; Heisig, Jan Paul; Schaeffer, Merlin; Shikano, Susumu

Article — Accepted Manuscript (Postprint)

Multilevel Analysis with Few Clusters: Improving Likelihood-based Methods to Provide Unbiased Estimates and Accurate Inference

British Journal of Political Science

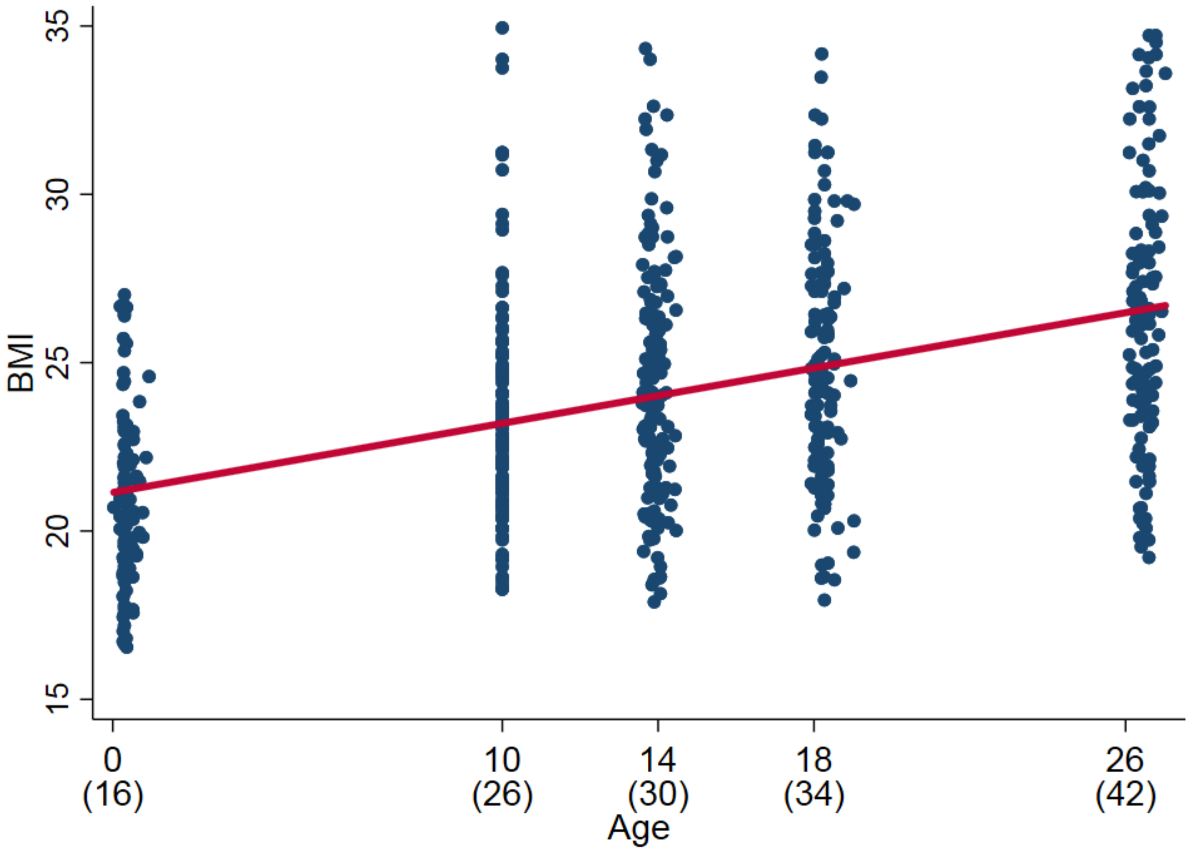
Regression



$$\text{bmi}_{ij} = \beta_{0i}\text{cons} + 0.205(0.016)\text{age}_{ij}$$

$$\beta_{0i} = 17.857(0.511) + e_{0ij}$$

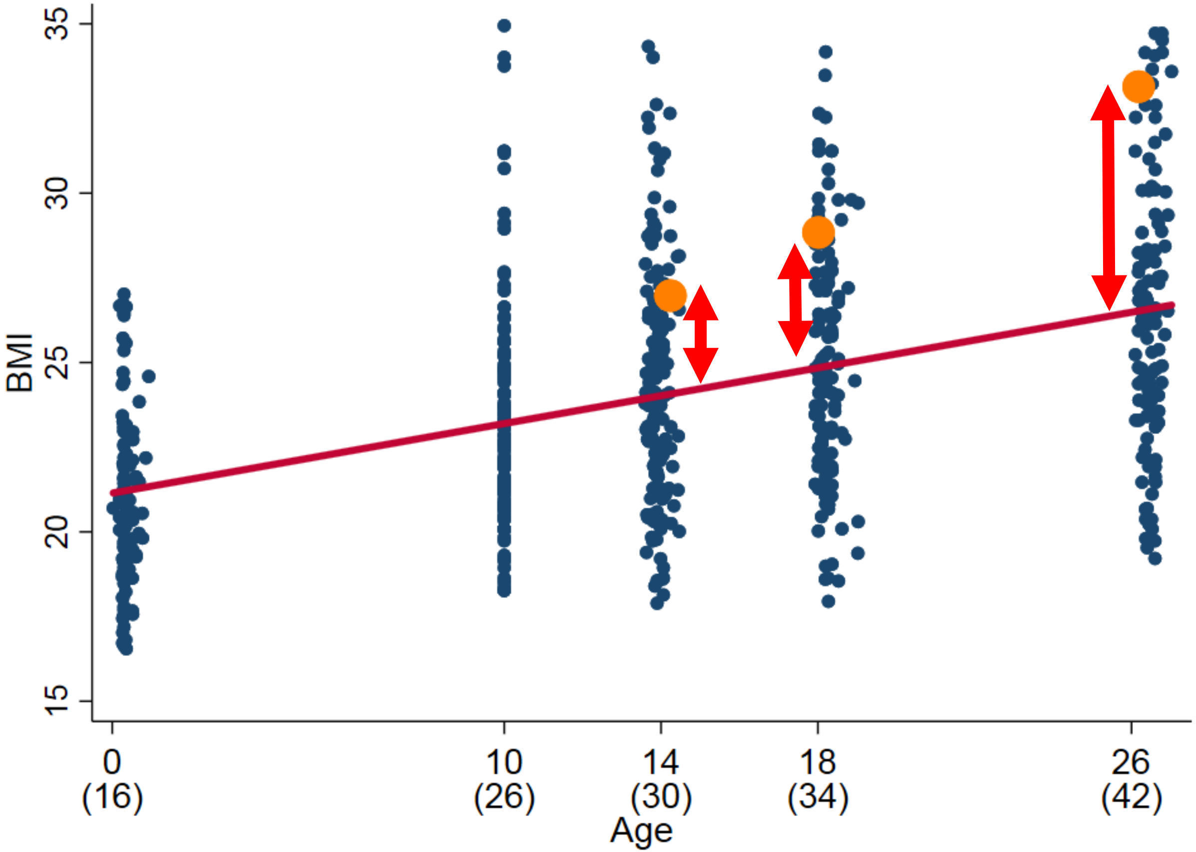
Regression



$$\text{bmi}_{ij} = \beta_{0i}\text{cons} + 0.205(0.016)\text{agec}_{ij}$$

$$\beta_{0i} = 21.144(0.267) + e_{0ij}$$

Regression

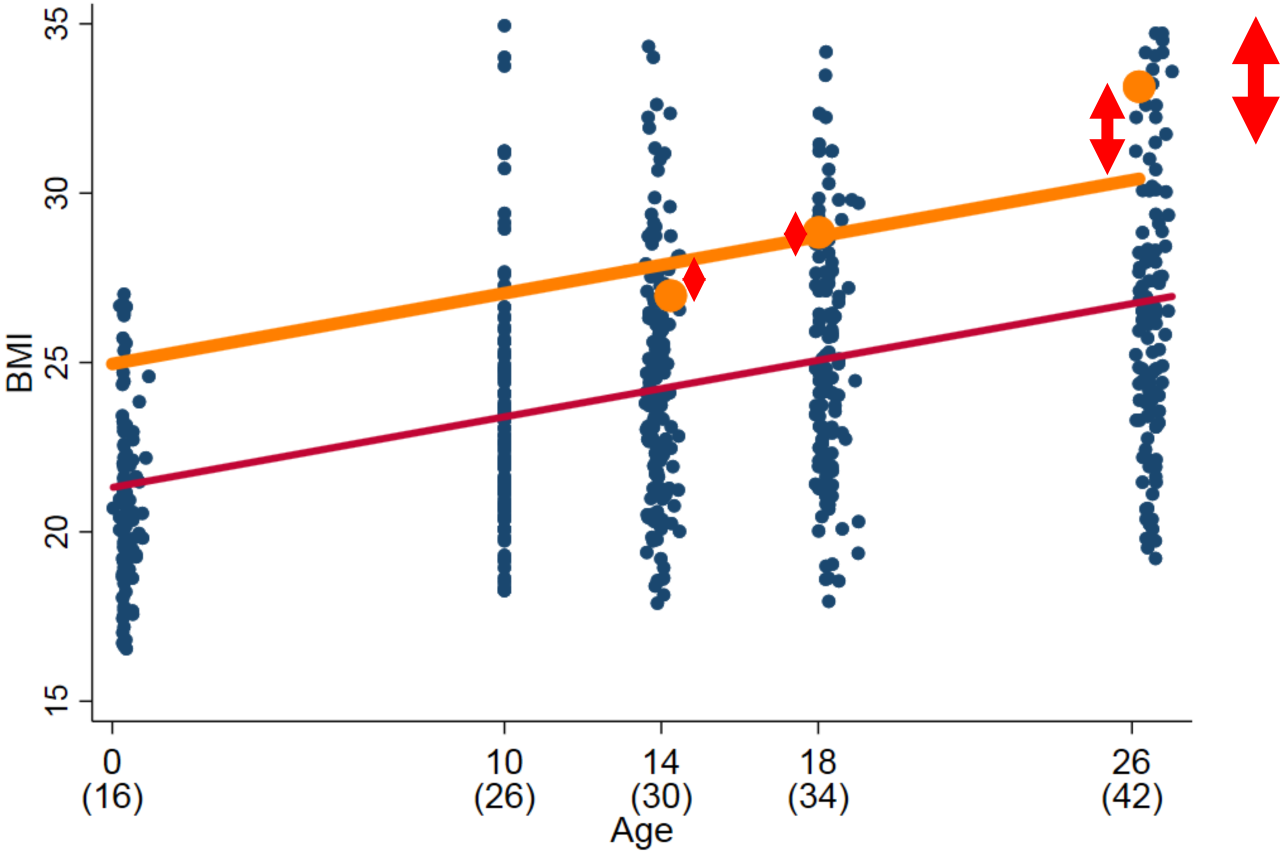


$$\text{bmi}_{ij} = \beta_{0i}\text{cons} + 0.205(0.016)\text{agec}_{ij}$$

$$\beta_{0i} = 21.144(0.267) + e_{0ij}$$

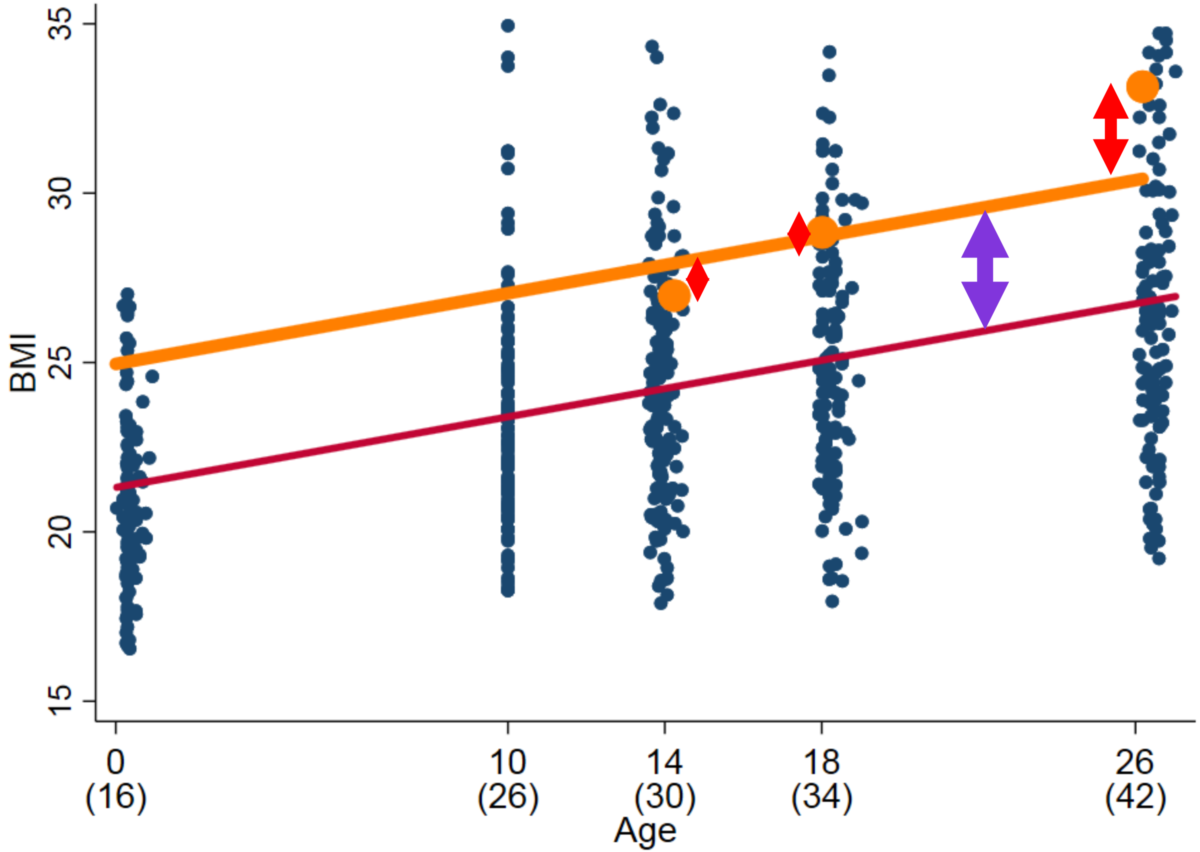
$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[11.674(0.661)\right]$$

Multilevel regression – random intercept



Level 1 or within-individual Residuals/ variance

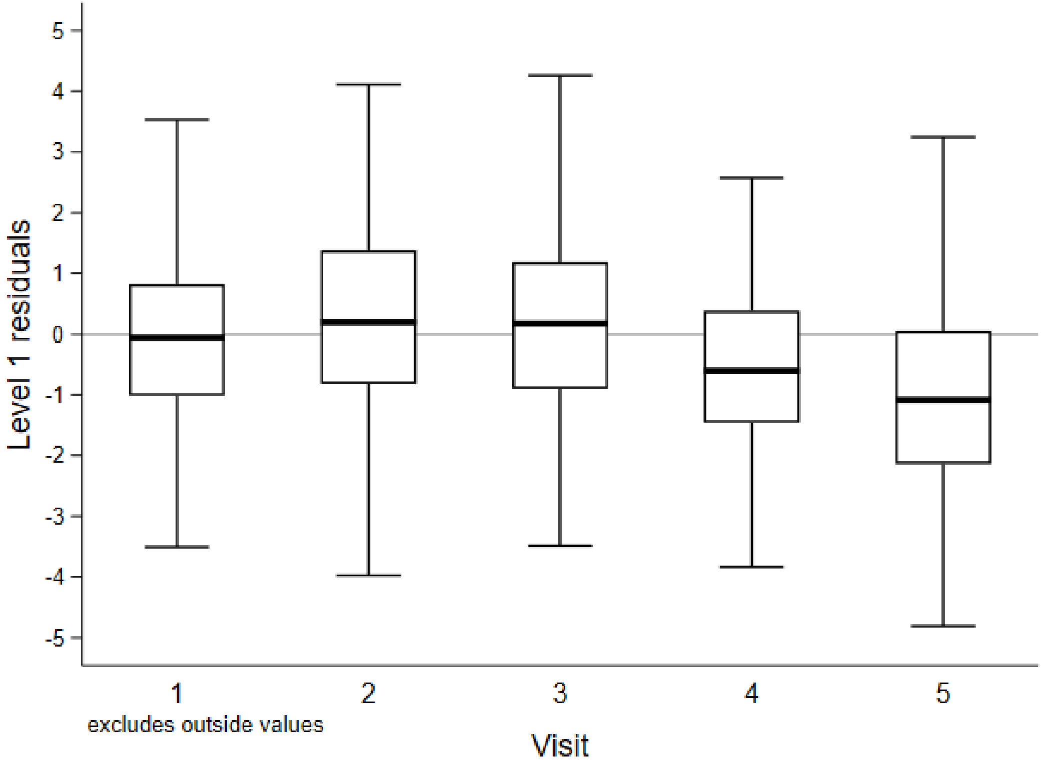
Multilevel regression – random intercept



Level 1 or within-individual Residuals/ variance

Level 2 or between-individual Residuals/ variance

Multilevel regression – random intercept



$$\text{bmi}_{ij} = \beta_{0ij}\text{cons} + \beta_1 \text{agec}_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u) : \Omega_u = \left[\sigma^2_{u0}\right]$$

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e) : \Omega_e = \left[\sigma^2_{e0}\right]$$

Multilevel regression – random intercept

$$\text{bmi}_{ij} = \beta_{0ij}\text{cons} + 0.209(0.011)\text{agec}_{ij}$$

$$\beta_{0ij} = 21.311(0.263) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[7.940(0.952)\right]$$

sqrt(7.940) = 2.8 kg/m^2

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[4.104(0.282)\right]$$

sqrt(4.104) = 2.0 kg/m^2

Multilevel regression – random intercept

$$\text{bmi}_{ij} = \beta_{0ij}\text{cons} + 0.209(0.011)\text{agec}_{ij}$$

$$\beta_{0ij} = 21.311(0.263) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[7.940(0.952)\right]$$

sqrt(7.940) = 2.8 kg/m^2

$$\left[e_{0ij}\right] \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[4.104(0.282)\right]$$

sqrt(4.104) = 2.0 kg/m^2

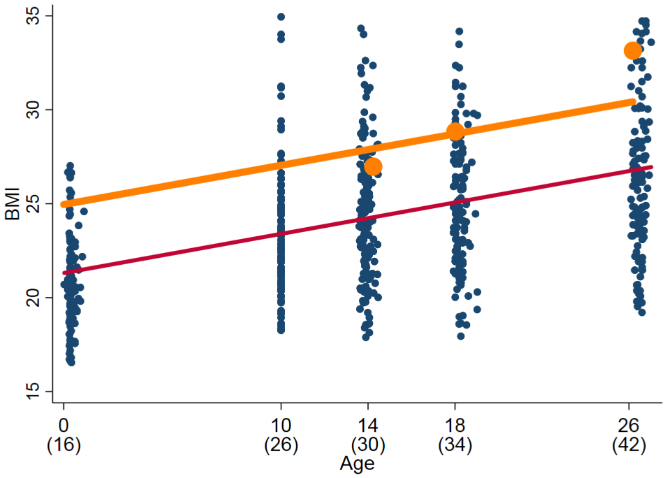
Variance partitioning coefficient

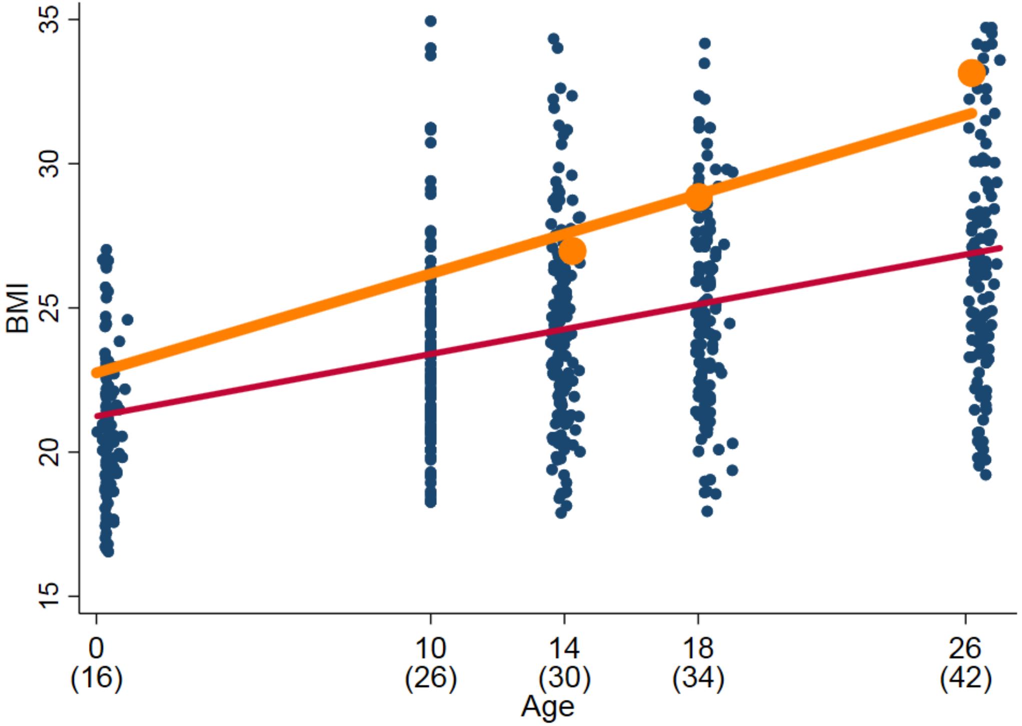
Proportion of total variance that is due to between-individual differences (i.e., level 2 variance)

= 7.940 / (4.103 + 7.940)

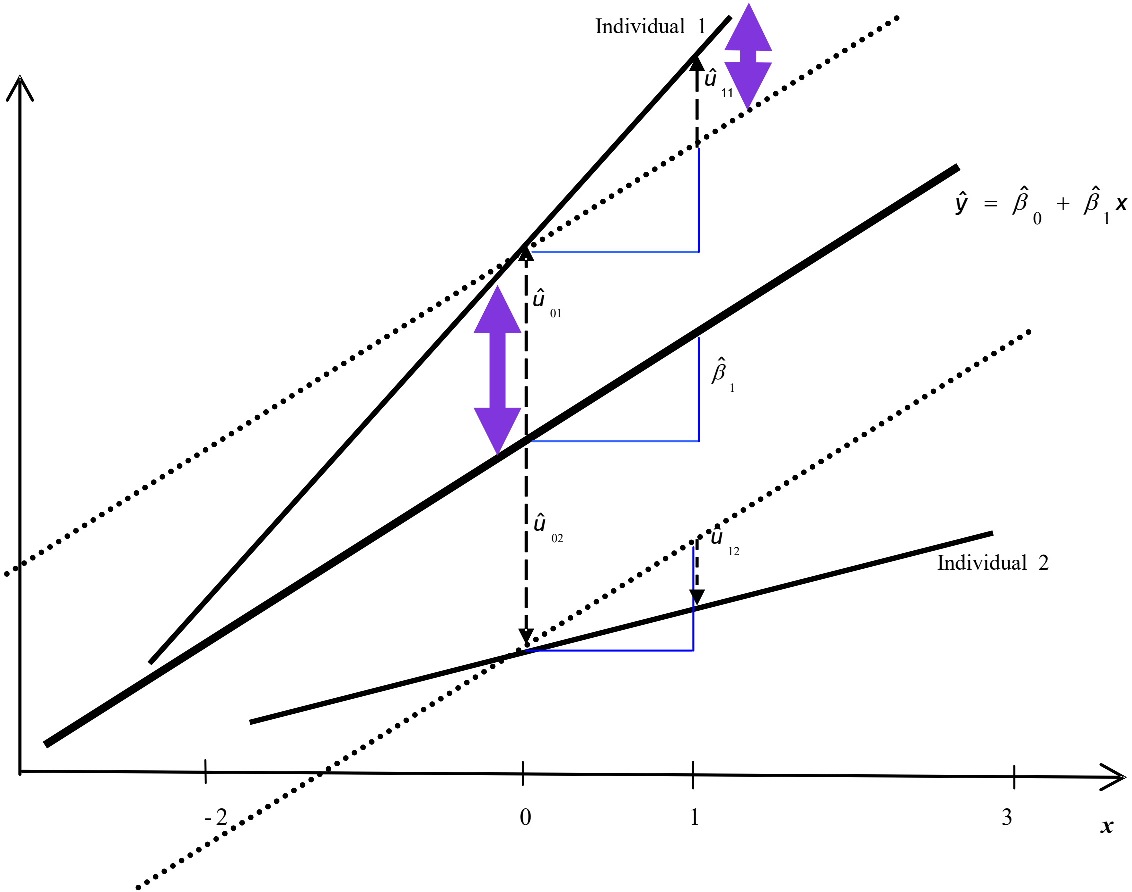
= 0.659

Multilevel regression – random slope





Multilevel regression – random slope



Multilevel regression – random slope

$$\text{bmi}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{agec}_{ij}$$

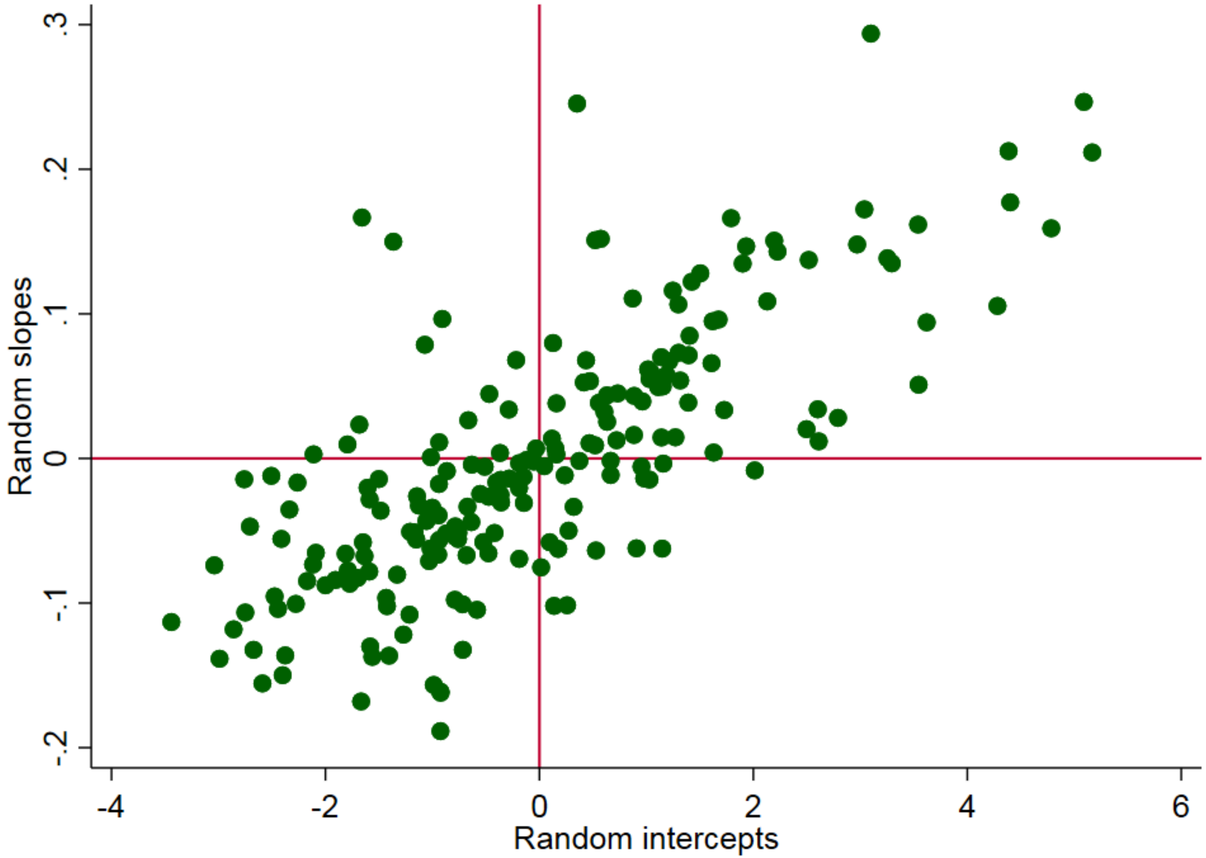
$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix} \qquad \begin{bmatrix} 4.283(0.862) & \\ 0.071(0.038) & 0.013(0.003) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix} \qquad \begin{bmatrix} 2.967(0.246) \end{bmatrix}$$

Multilevel regression – random slope

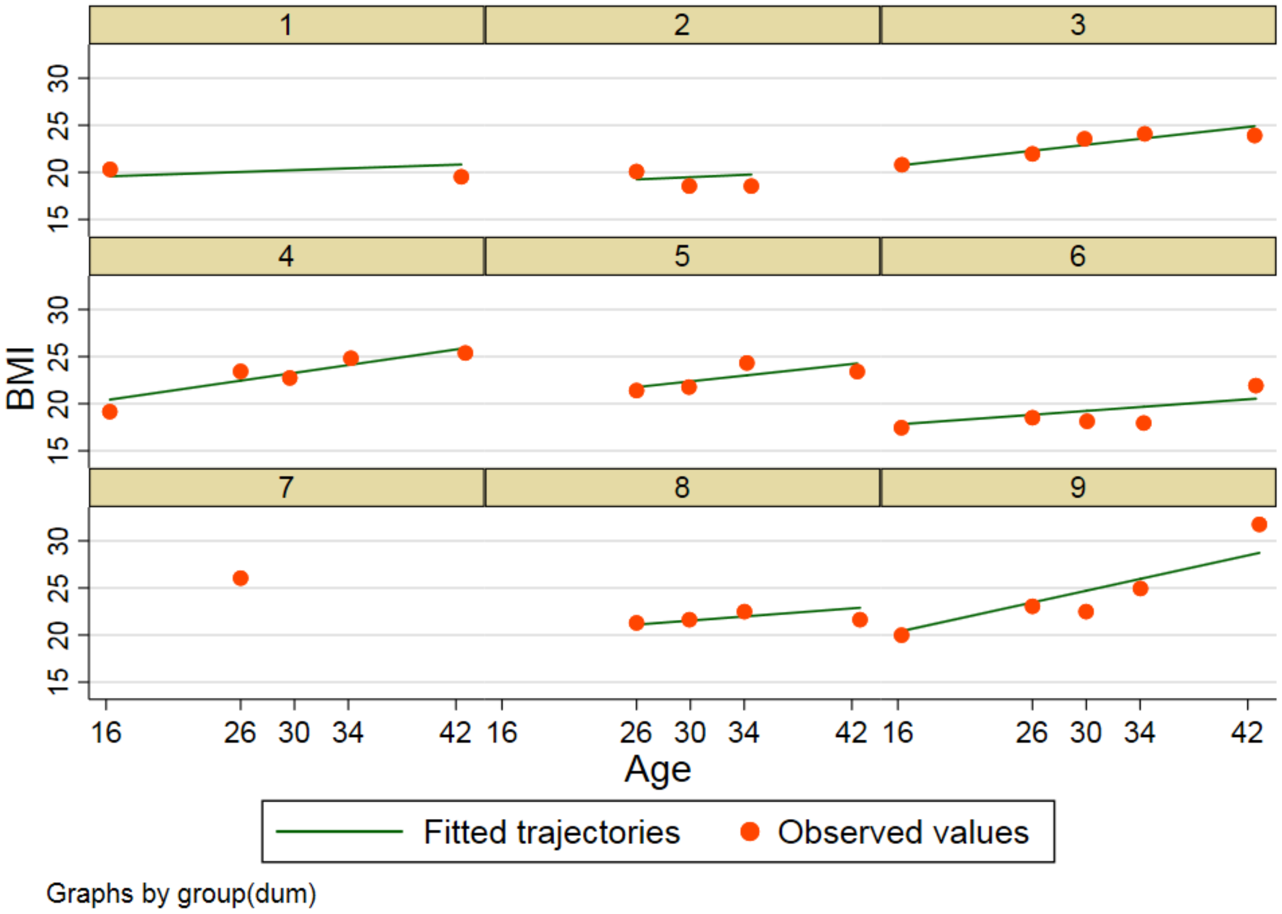


Covariance
0.071

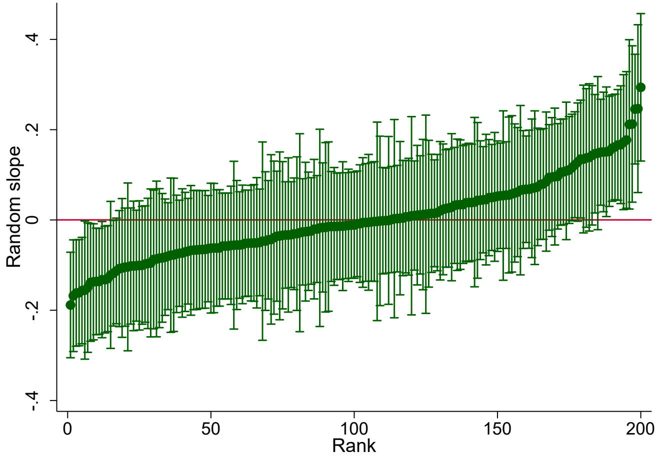
=

Correlation
0.298

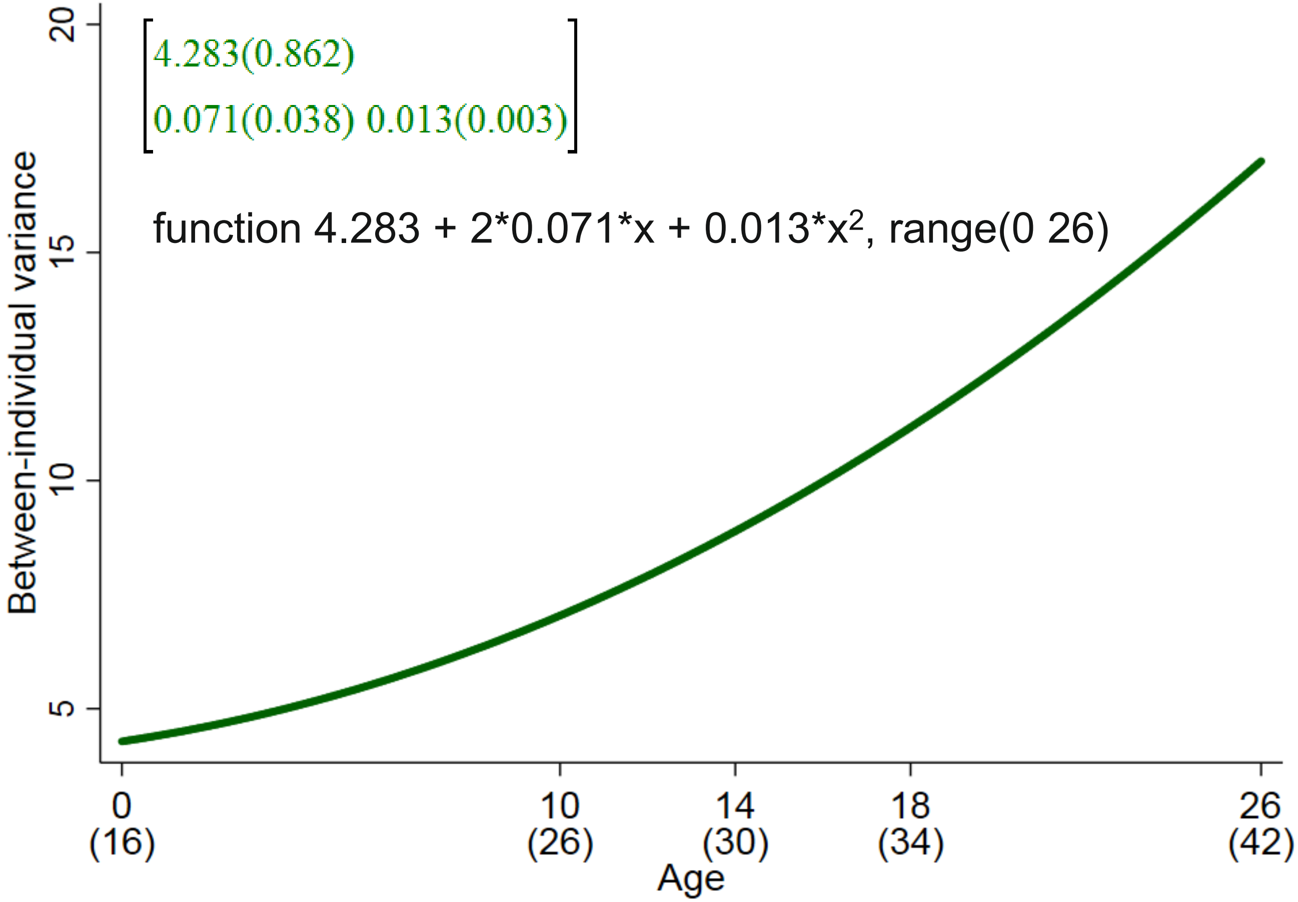
Multilevel regression – random slope



Multilevel regression – random slope

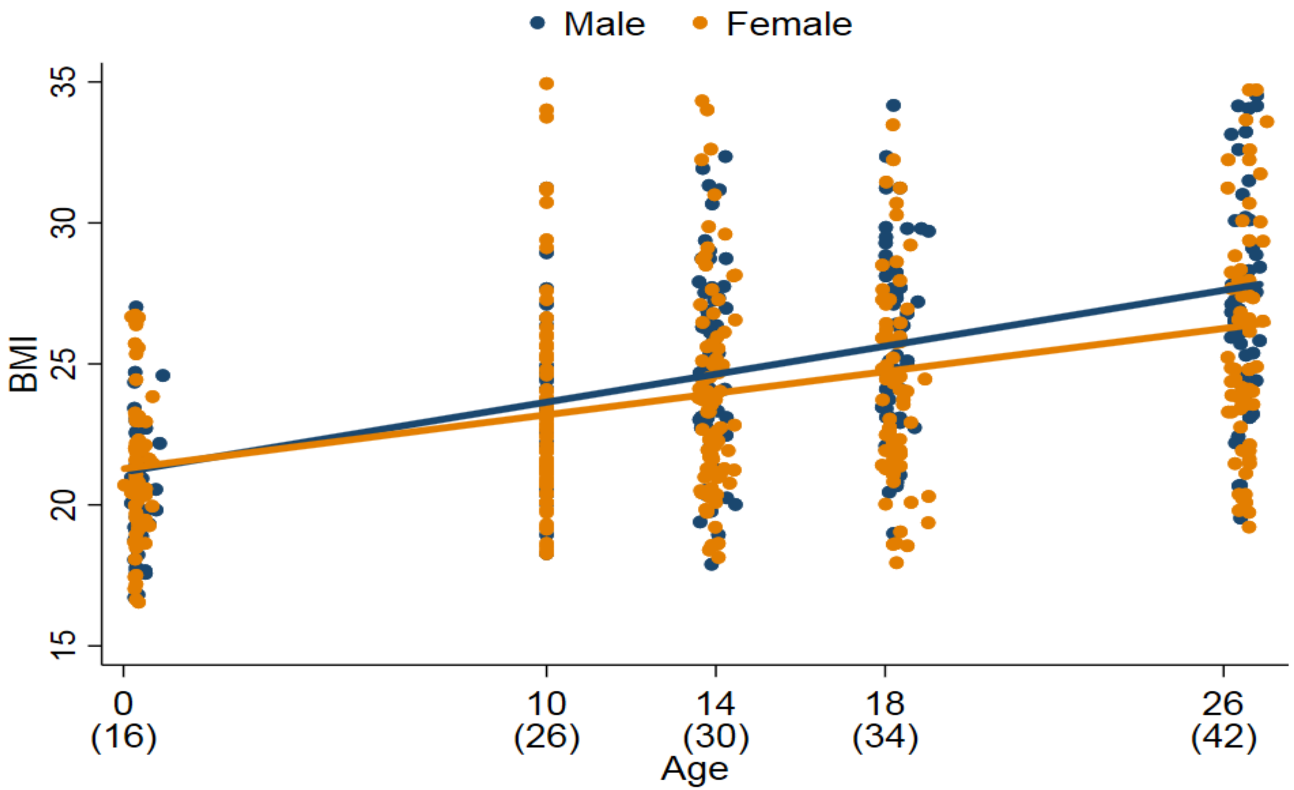


Multilevel regression – random slope



Including explanatory variables – level 2

$$\text{bmi}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{agec}_{ij} + 0.118(0.427)\text{sex}_j + -0.057(0.026)\text{agec\_sex}_{ij}$$



Including explanatory variables – level 1

$$\text{bmi}_{ij} = \beta_{0j}\text{cons} + \beta_{1j}\text{agec}_{ij} + 0.115(0.423)\text{sex}_j + -0.057(0.025)\text{agec\_sex}_{ij} + e_{4ij}\text{male}_j + e_{5ij}\text{female}_j$$

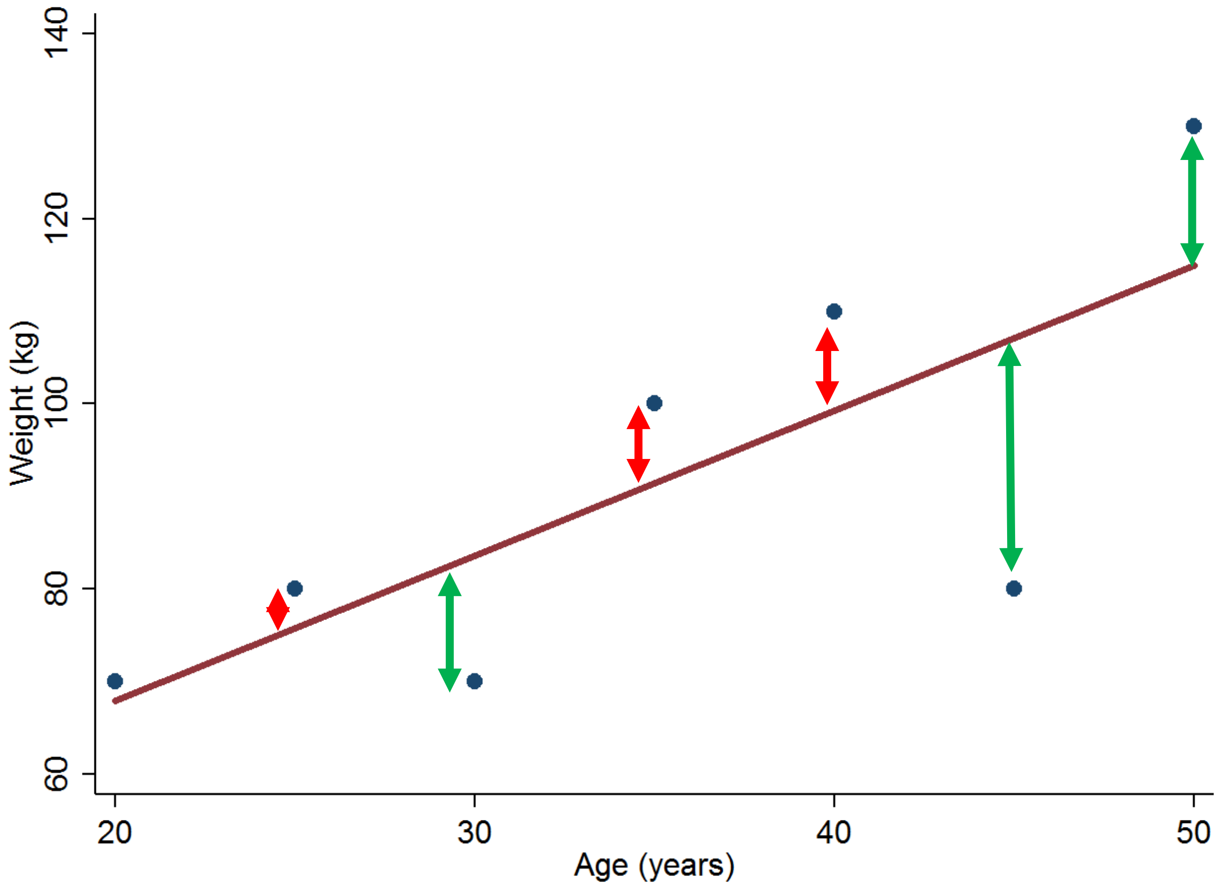
$$\beta_{0j} = 21.171(0.319) + u_{0j}$$

$$\beta_{1j} = 0.248(0.019) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} 4.189(0.851) & \\ 0.074(0.037) & 0.012(0.003) \end{bmatrix}$$

$$\begin{bmatrix} e_{4ij} \\ e_{5ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} 2.699(0.343) & \\ 0 & 3.202(0.322) \end{bmatrix}$$

Including explanatory variables – level 1



Measured

Self-reported

Recap

With longitudinal data:

- Multilevel models account for clustering by partitioning variance into within-individual (level 1) and between-individual (level 2)
- A separate trajectory, describing how y changes over age/time, is essentially estimated for each individual
    - Fixed intercept/slope = sample average trajectory
    - Fixed intercept/slope + random intercept/slope = individual trajectories
- Additional explanatory variables at level 1 (self-report) or 2 (sex) can be included to model systematic differences between-individuals
- Additional explanatory variables at level 1 (self-report) or 2 (sex) can be included to model systematic differences within-individuals

Does accounting for clustering really matter?

Fixed intercept & slope

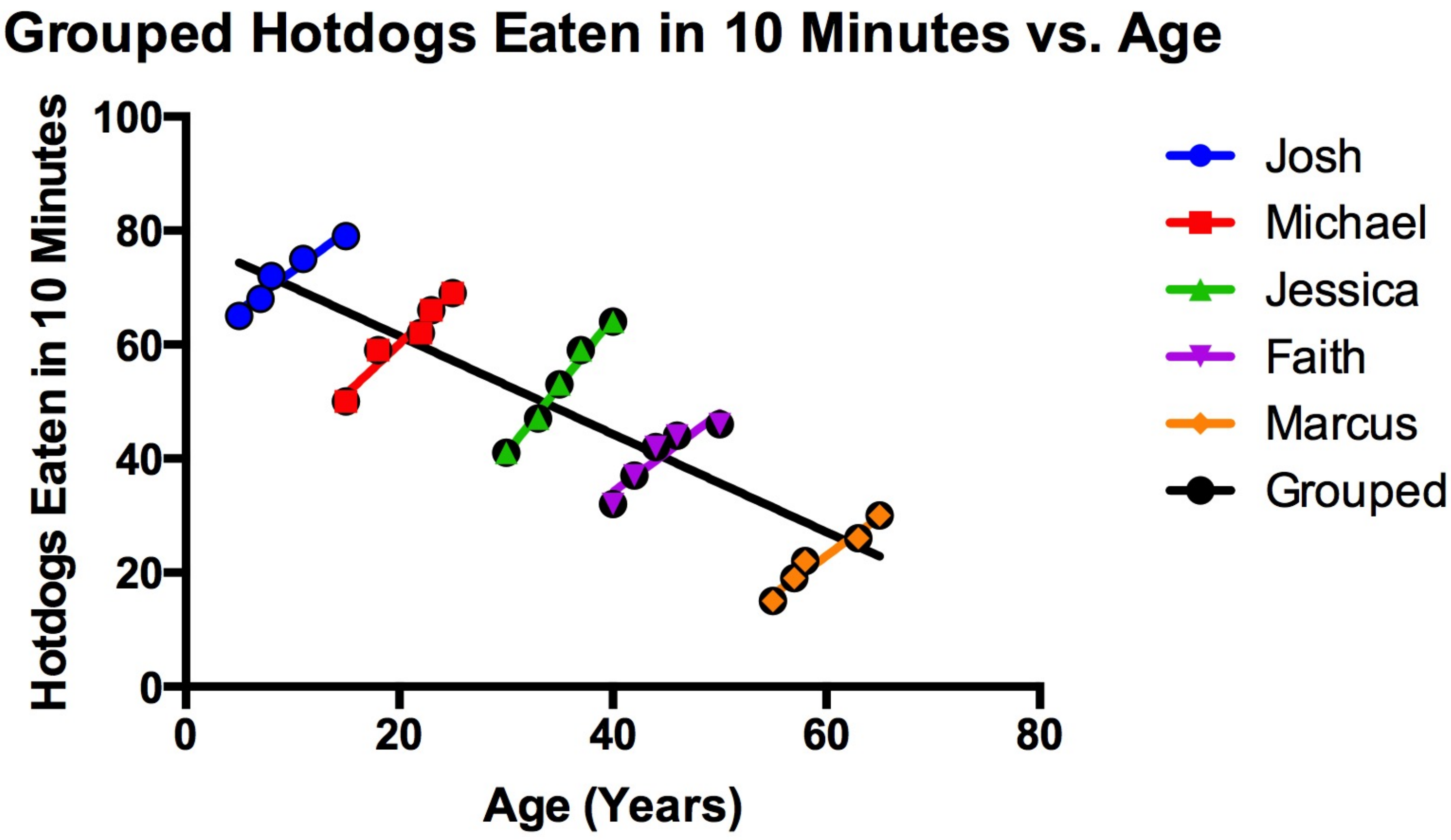
$\beta_1 = 0.205(0.016)$

Vs.

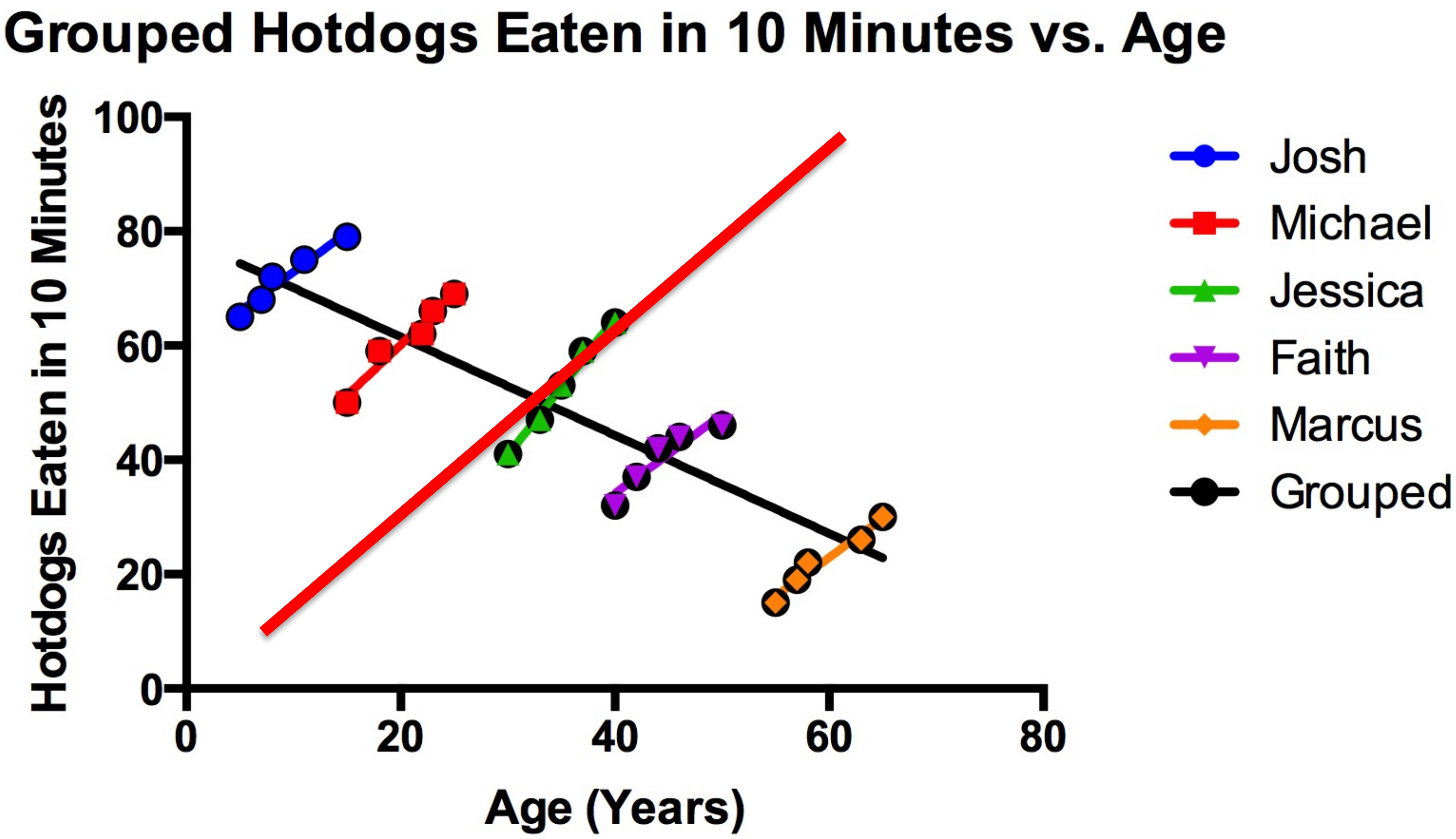
Random intercept & slope

$\beta_1 = 0.216(0.013)$

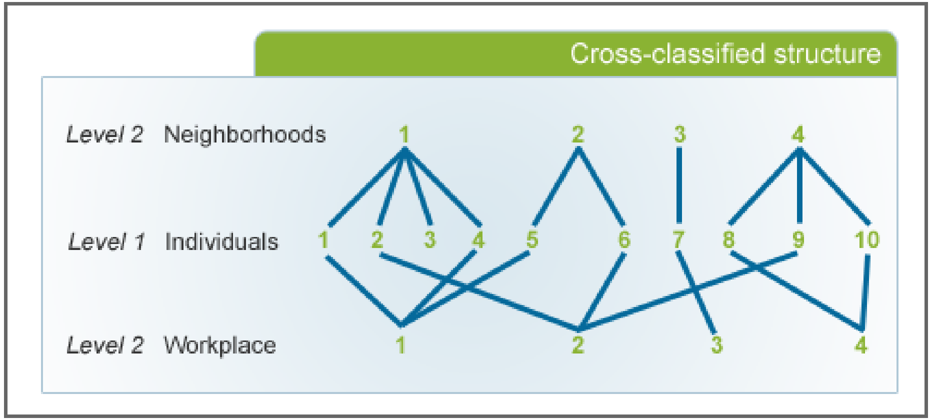
Does accounting for clustering really matter?

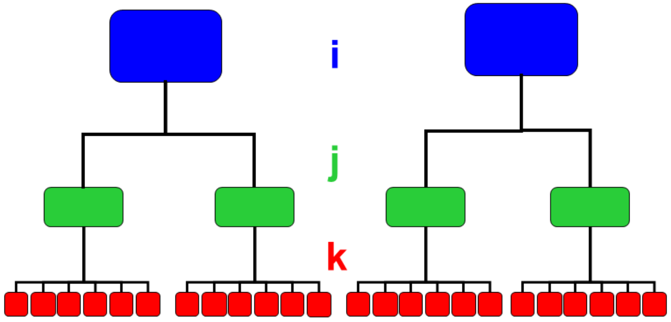


Does accounting for clustering really matter?



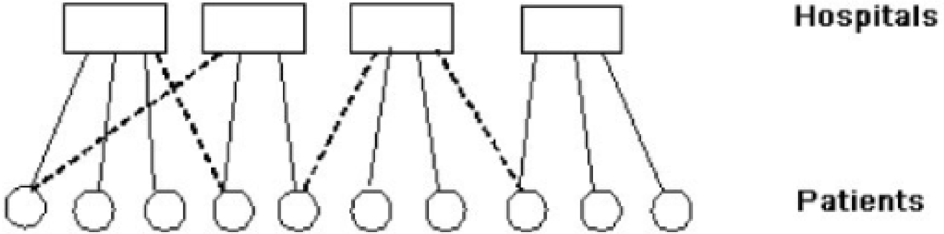
Other structures





$$y_{ijk} = y_{region,state,observation}$$

Multiple membership model: some patients attend more than one hospital



Other structures

International Journal of
Environmental Research and Public Health

Article

A Multilevel Analysis of Neighbourhood, School, Friend and Individual-Level Variation in Primary School Children's Physical Activity

Ruth Salway [1], Lydia Emm-Collison [1], Simon J. Sebire [1], Janice L. Thompson [2], Deborah A. Lawlor [3,4] and Russell Jago [1,*]

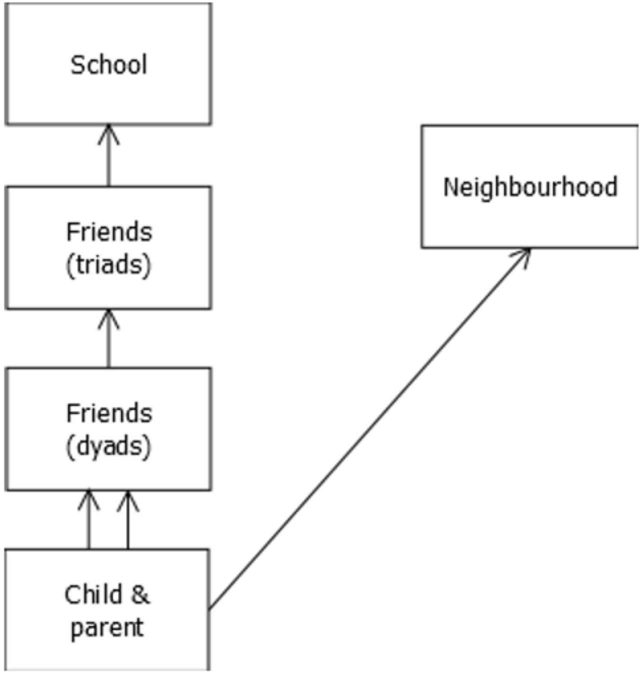
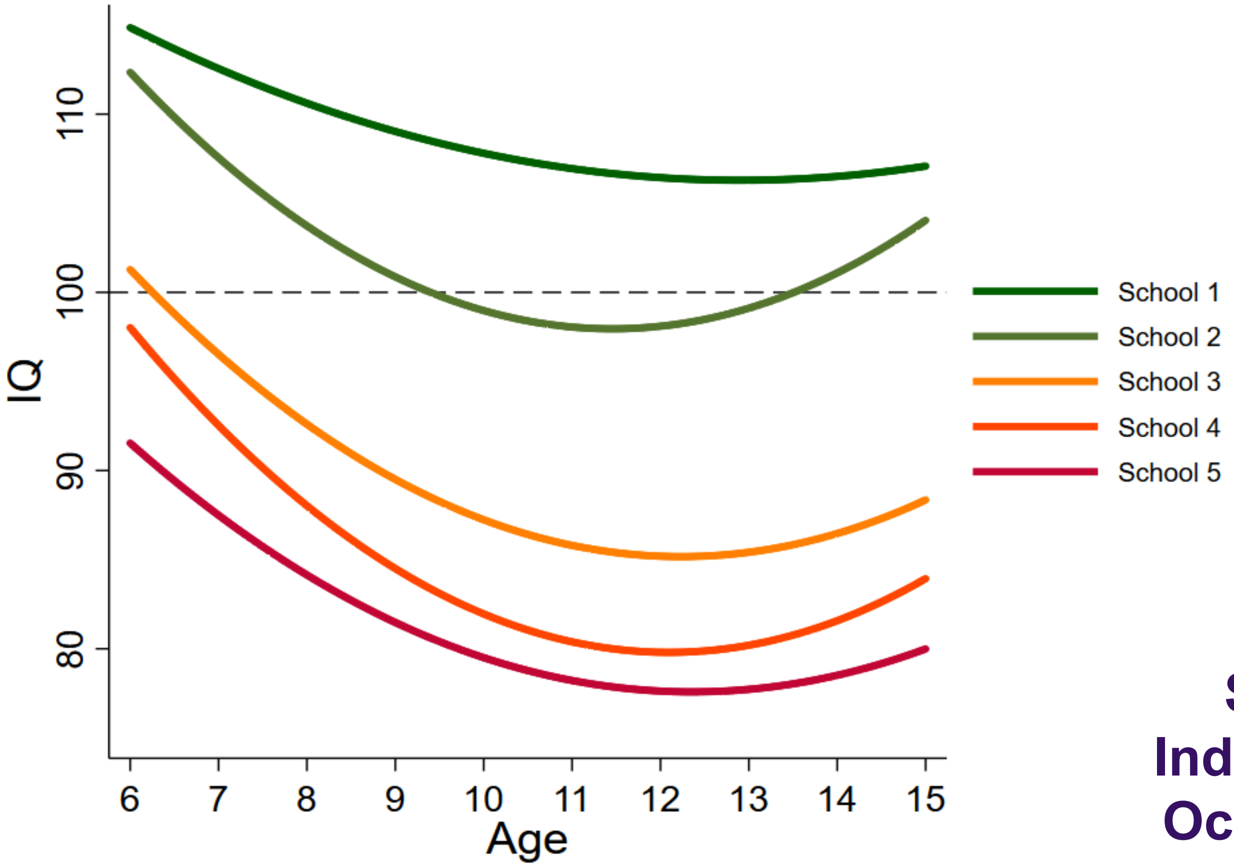


Figure 1. Classification diagram for the multiple-membership multiple-classification model. An arrow indicates a nested relationship; a double arrow indicates multiple membership.

Non-linear trajectories



Schools (level 3)
Individuals (level 2)
Occasions (level 1)

Non-linear trajectories

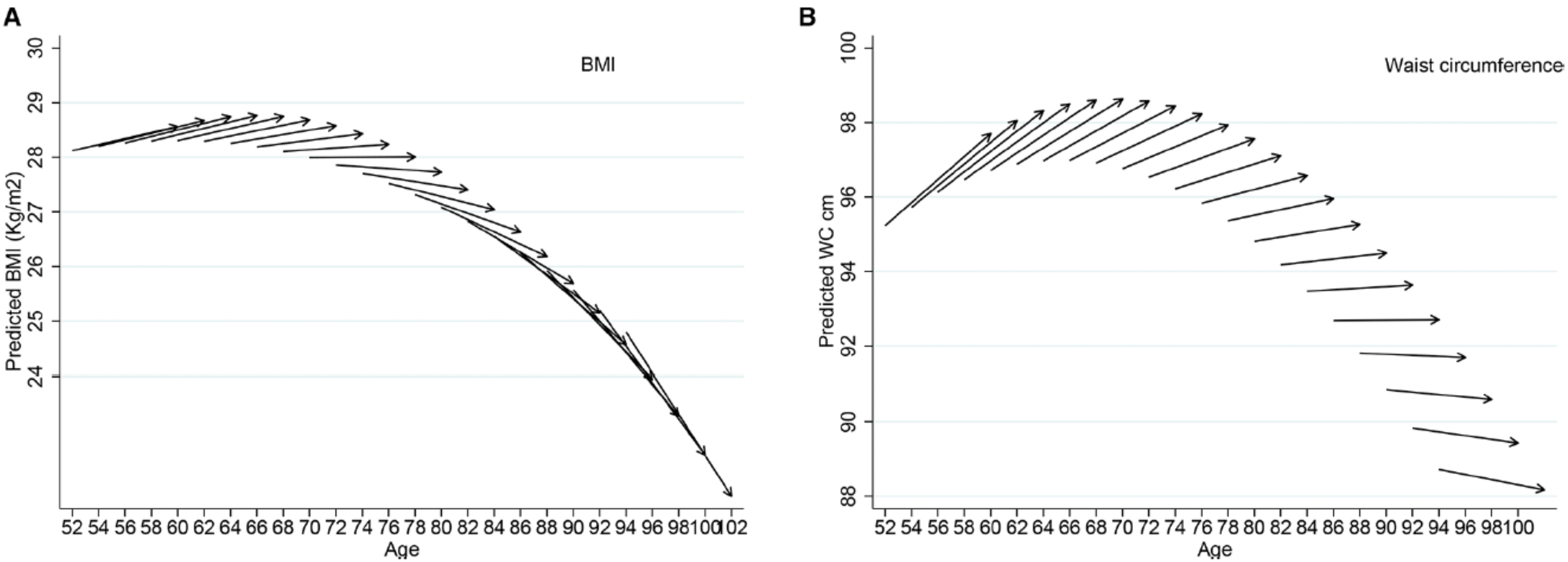
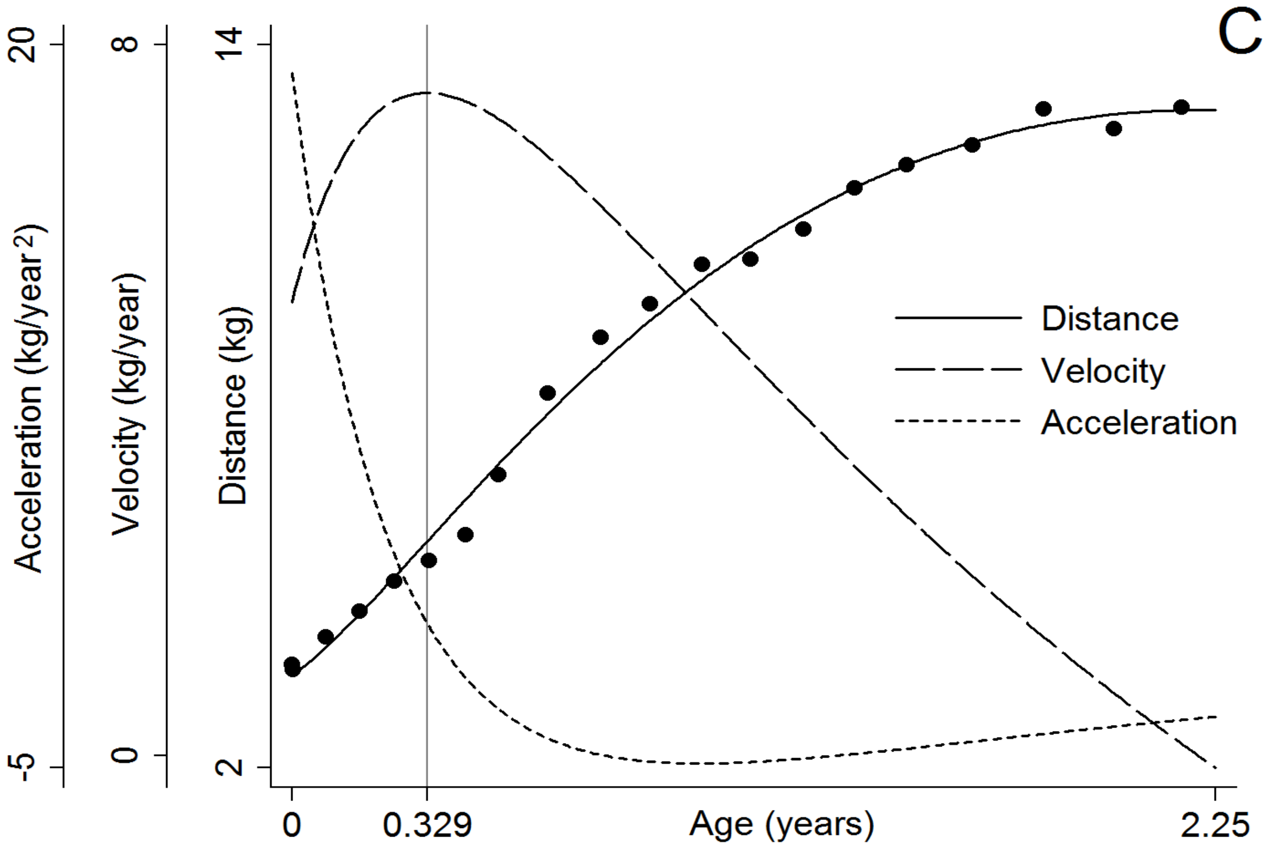


Figure 1 Vector graph showing 8-year ageing vectors of anthropometric markers, ELSA 2004–2005 to 2012–2013. BMI, body mass index; ELSA, English Longitudinal Study of Ageing; WC, waist circumference.

Zaninotto P, Lassale C. *BMJ Open* 2019;**9**:e025309. doi:10.1136/bmjopen-2018-025309

Derivatives and traits



2015. Am J Hum Biol 27(1): 69-83

Derivatives and traits

Journal of Youth and Adolescence (2019) 48:815–827
https://doi.org/10.1007/s10964-018-0976-5

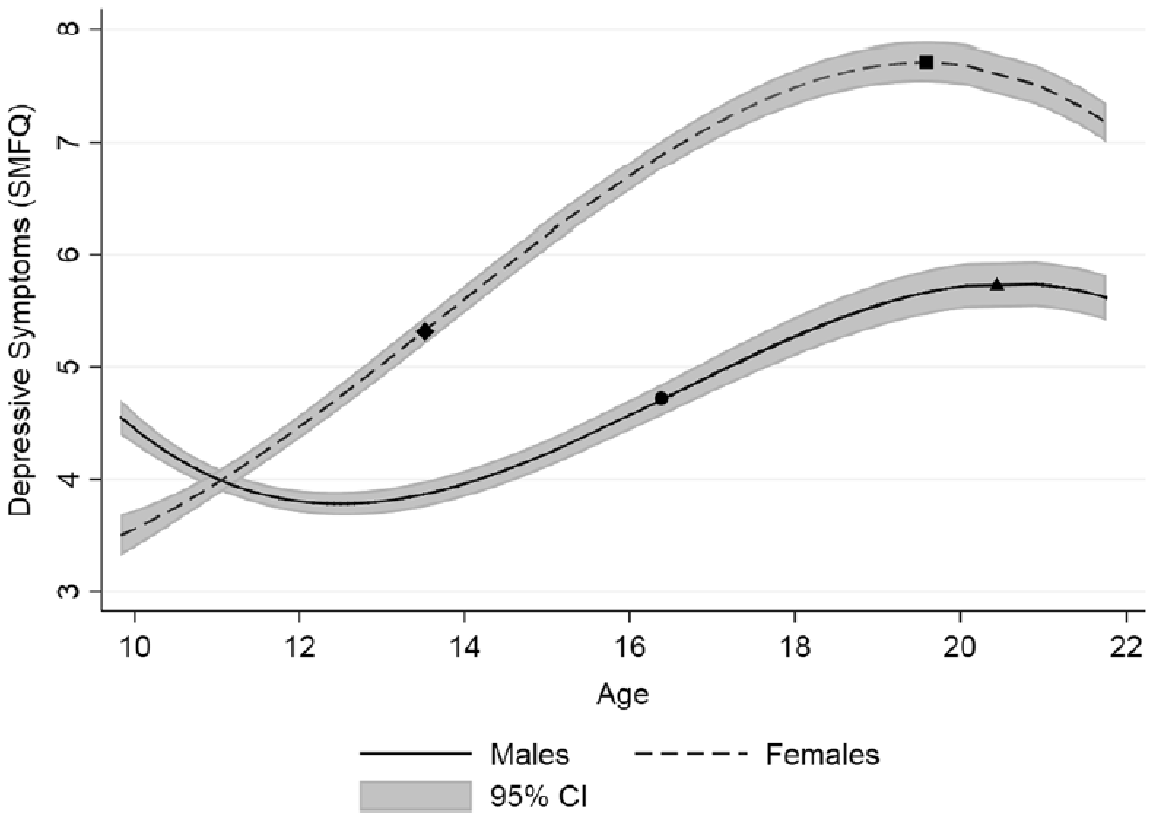
EMPIRICAL RESEARCH

CrossMark

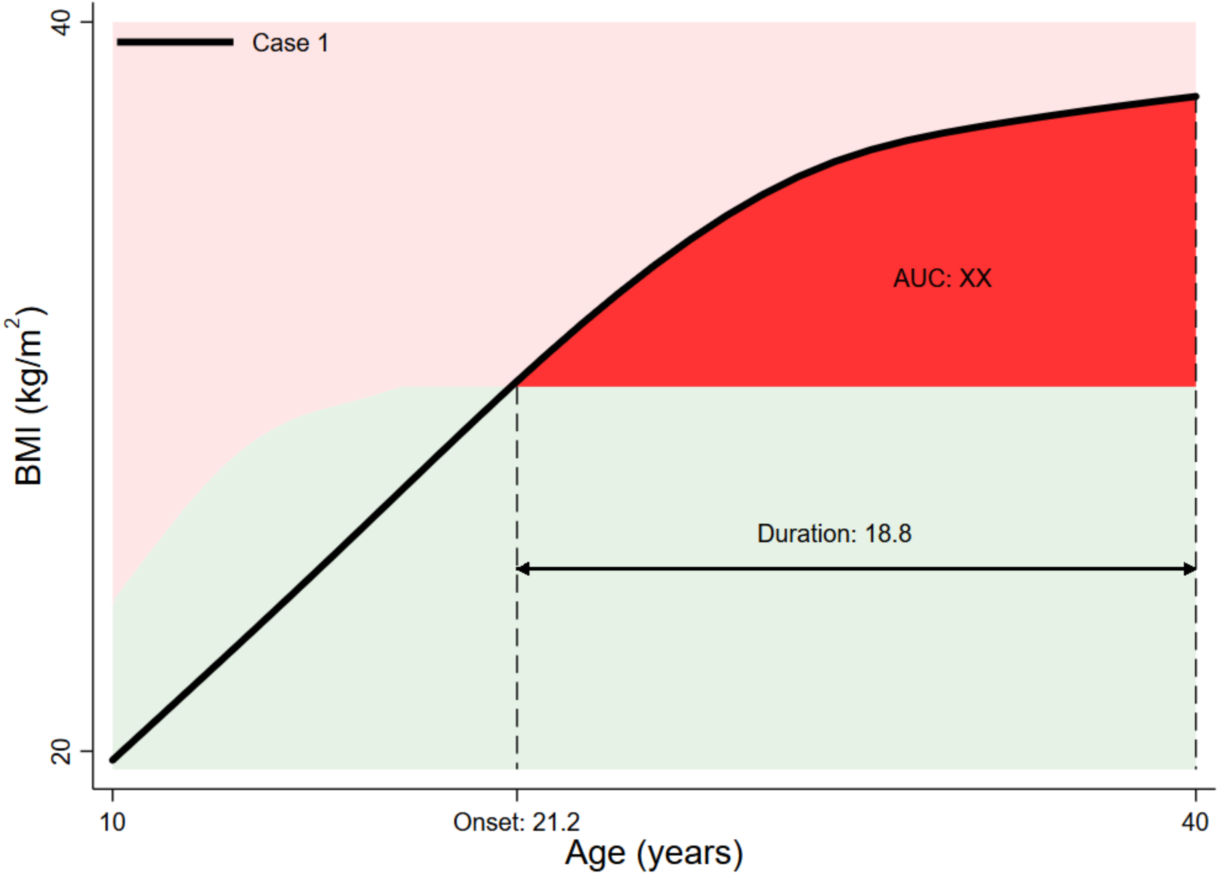
Identifying Critical Points of Trajectories of Depressive Symptoms from Childhood to Young Adulthood

Alex S. F. Kwong[1,2,3] · David Manley[1,2] · Nicholas J. Timpson[3,4] · Rebecca M. Pearson[3,4,5] · Jon Heron[3,4,5] · Hannah Sallis[3,4,5,6] · Evie Stergiakouli[3,4,7] · Oliver S. P. Davis[3,4] · George Leckie[2,8]

Received: 7 December 2018 / Accepted: 8 December 2018 / Published online: 22 January 2019



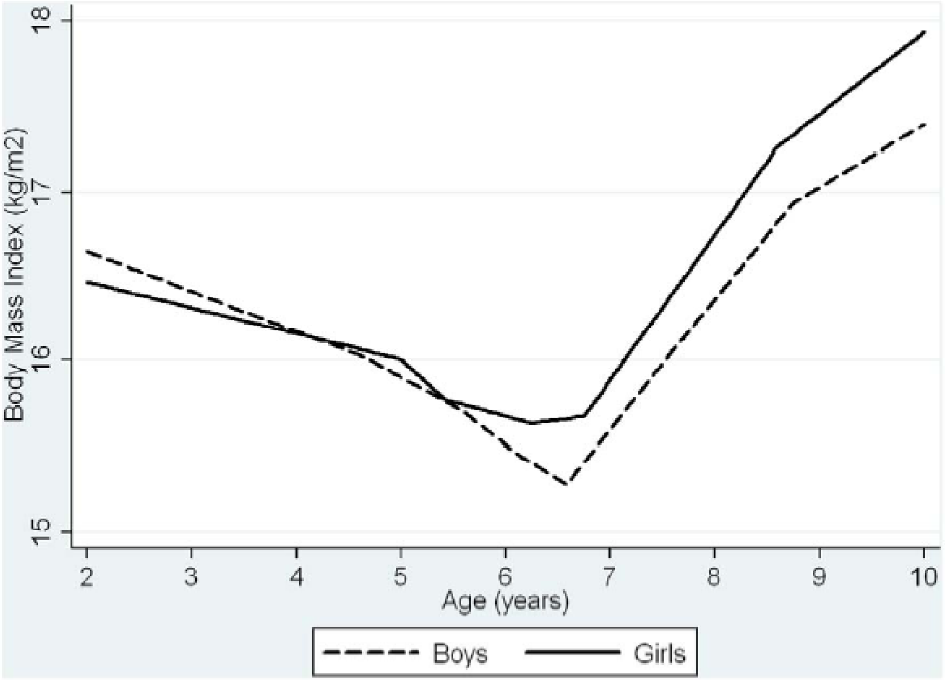
Derivatives and traits



Relating level 2 residuals to distal outcomes

Changes in Ponderal Index and Body Mass Index across Childhood and Their Associations with Fat Mass and Cardiovascular Risk Factors at Age 15

Laura D. Howe[1,2]*, Kate Tilling[2], Li Benfield[1,2], Jennifer Logue[3], Naveed Sattar[3], Andy R. Ness[4], George Davey Smith[1,2], Debbie A. Lawlor[1,2]



Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2016, Vol. 25(5) 1854–1874

DOI: 10.1177/0962280213503925
smm.sagepub.com
SAGE

Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts

Laura D Howe,[1] Kate Tilling,[2] Alicia Matijasevich,[3] Emily S Petherick,[4] Ana Cristina Santos,[5] Lesley Fairley,[4] John Wright,[4] Iná S Santos,[3] Aluísio JD Barros,[3] Richard M Martin,[2,6] Michael S Kramer,[7] Natalia Bogdanovich,[8] Lidia Matush,[8] Henrique Barros[5] and Debbie A Lawlor[1]

Relating level 2 residuals to distal outcomes

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2017, Vol. 26(1) 437–452

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280214548822
smm.sagepub.com
SAGE

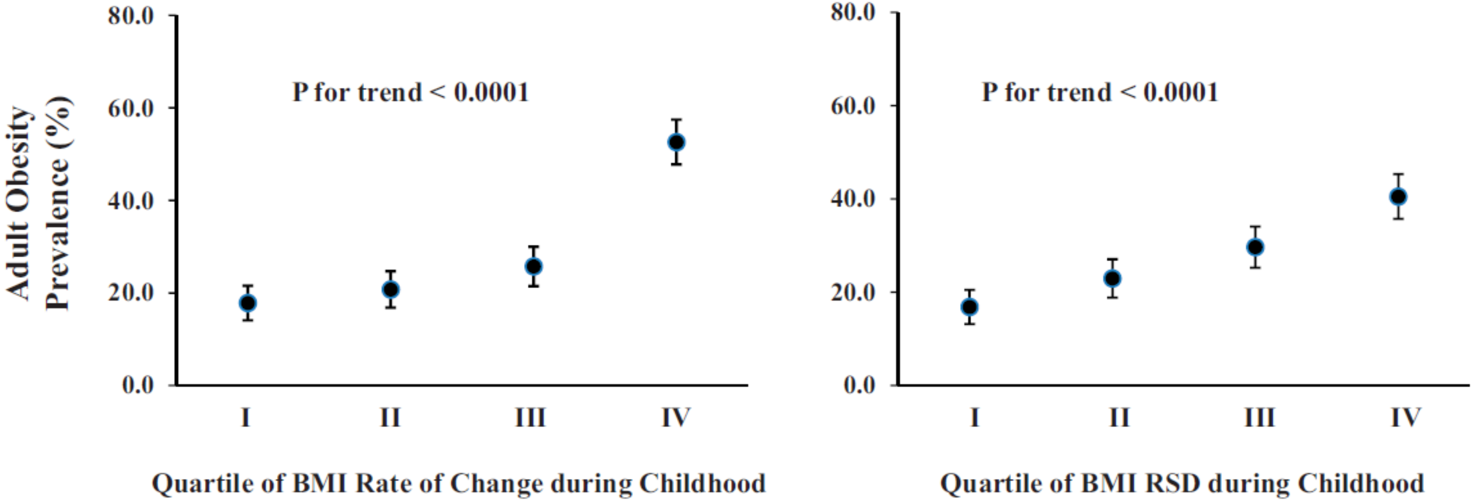
Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP

A Sayers,[1] J Heron,[1] ADAC Smith,[1,2] C Macdonald-Wallis,[1,2] MS Gilthorpe,[3] F Steele[4] and K Tilling[1,2]

Relating level 1 residuals to distal outcomes

Variability and rapid increase in body mass index during childhood are associated with adult obesity

Shengxu Li,[1]* Wei Chen,[1] Dianjianyi Sun,[1,2] Camilo Fernandez,[1] Jian Li,[3] Tanika Kelly,[1] Jiang He,[1] Marie Krousel-Wood[1,4,5] and Paul K Whelton[1]



Relating level 1 residuals to distal outcomes

RESEARCH ARTICLE

WILEY Statistics in Medicine

Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study

Jessica K. Barrett[1,2] | Raphael Huille[2,3] | Richard Parker[4] | Yuichiro Yano[5] | Michael Griswold[6]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[3]École Nationale de la Statistique et de l'Administration Économique, Malakoff, France

[4]School of Social and Community Medicine, University of Bristol, Bristol, UK

[5]Department of Preventive Medicine, University of Mississippi Medical Center, Jackson, Mississippi

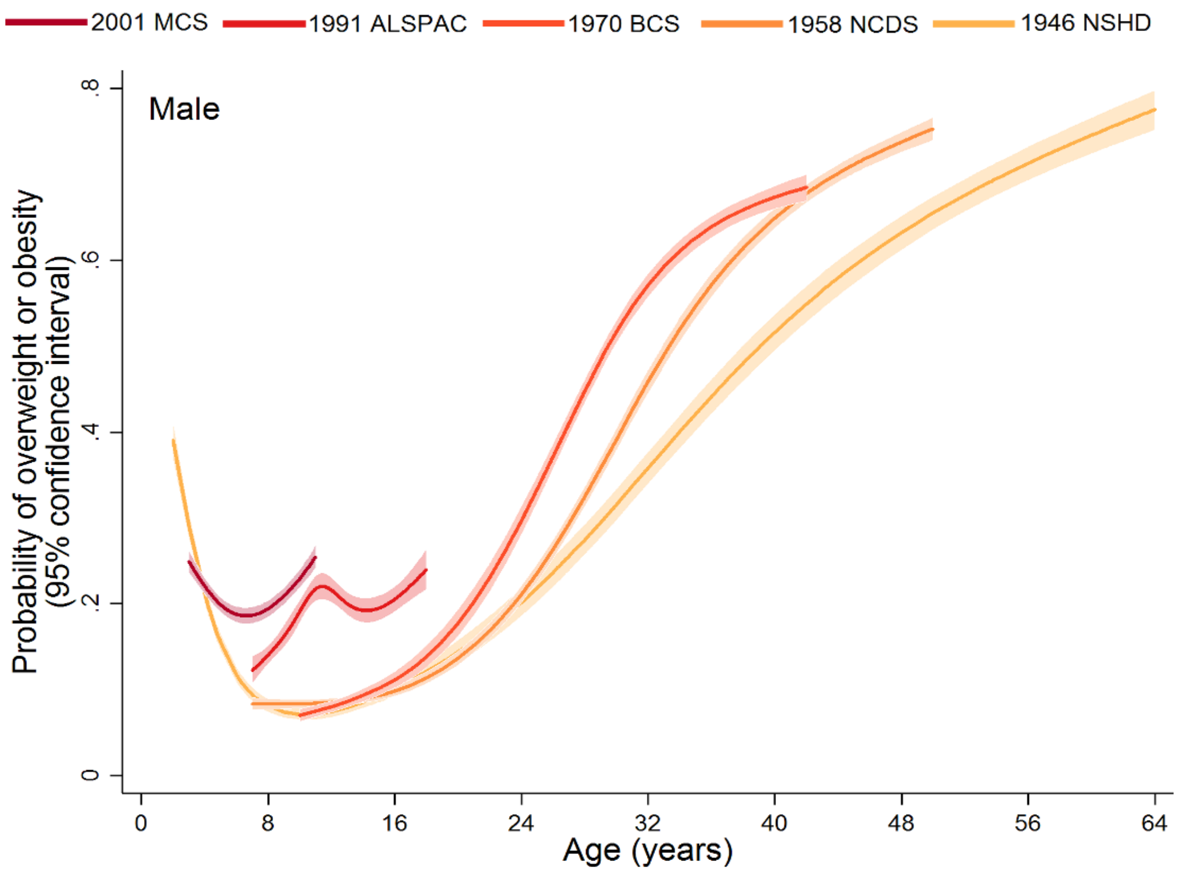
[6]Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, Mississippi

Correspondence

Jessica K. Barrett, MRC Biostatistics Unit,

The association between visit-to-visit systolic blood pressure variability and cardiovascular events has recently received a lot of attention in the cardiovascular literature. But, blood pressure variability is usually estimated on a person-by-person basis and is therefore subject to considerable measurement error. We demonstrate that hazard ratios estimated using this approach are subject to bias due to regression dilution, and we propose alternative methods to reduce this bias: a two-stage method and a joint model. For the two-stage method, in stage one, repeated measurements are modelled using a mixed effects model with a random component on the residual standard deviation (SD). The mixed effects model is used to estimate the blood pressure SD for each individual, which, in stage two, is used as a covariate in a time-to-event model. For the joint model, the mixed effects submodel and time-to-event submodel are fitted simultaneously using shared random effects. We illustrate the methods using data from the Atherosclerosis Risk in Communities study.

Other outcome distributions



2015. Plos Med 12: e1001828

Parallel processes

Research Article

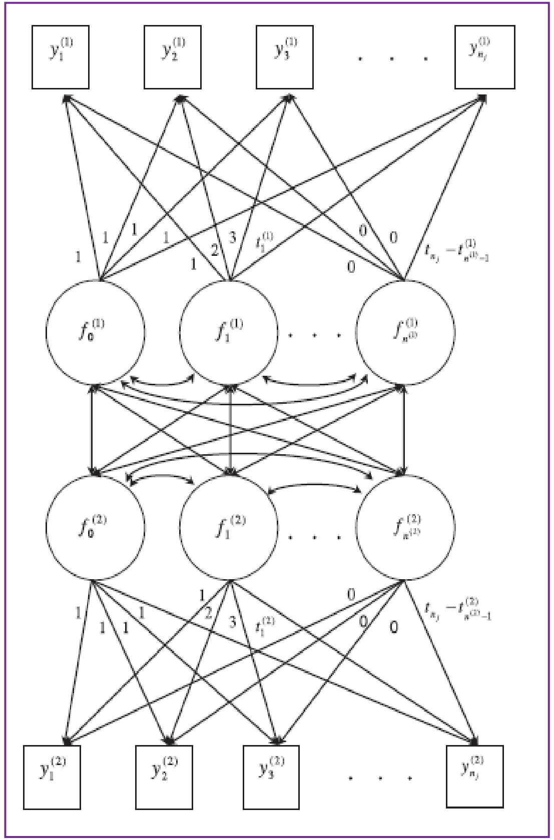
Statistics in Medicine

Received 8 March 2011, Accepted 6 March 2012 Published online 26 June 2012 in Wiley Online Library

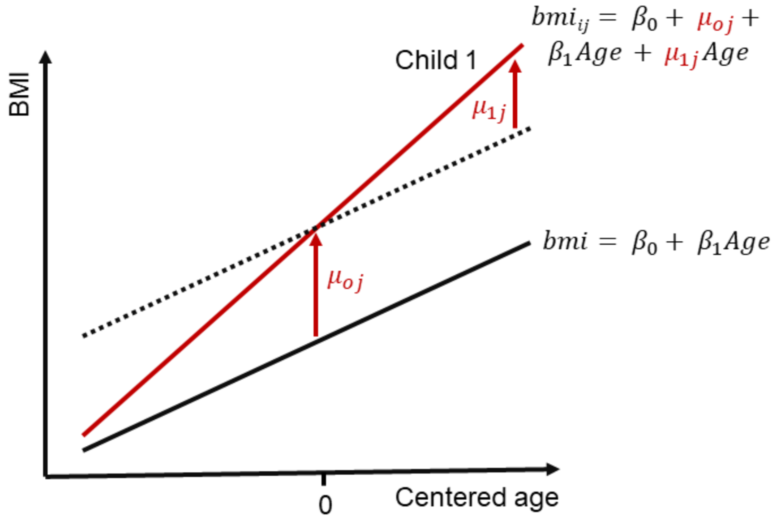
(wileyonlinelibrary.com) DOI: 10.1002/sim.5385

Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy

Corrie Macdonald-Wallis,[a*†] Debbie A. Lawlor,[a] Tom Palmer[a] and Kate Tilling[b]

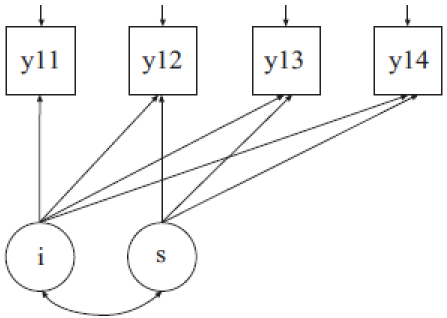


Extend to SEM framework



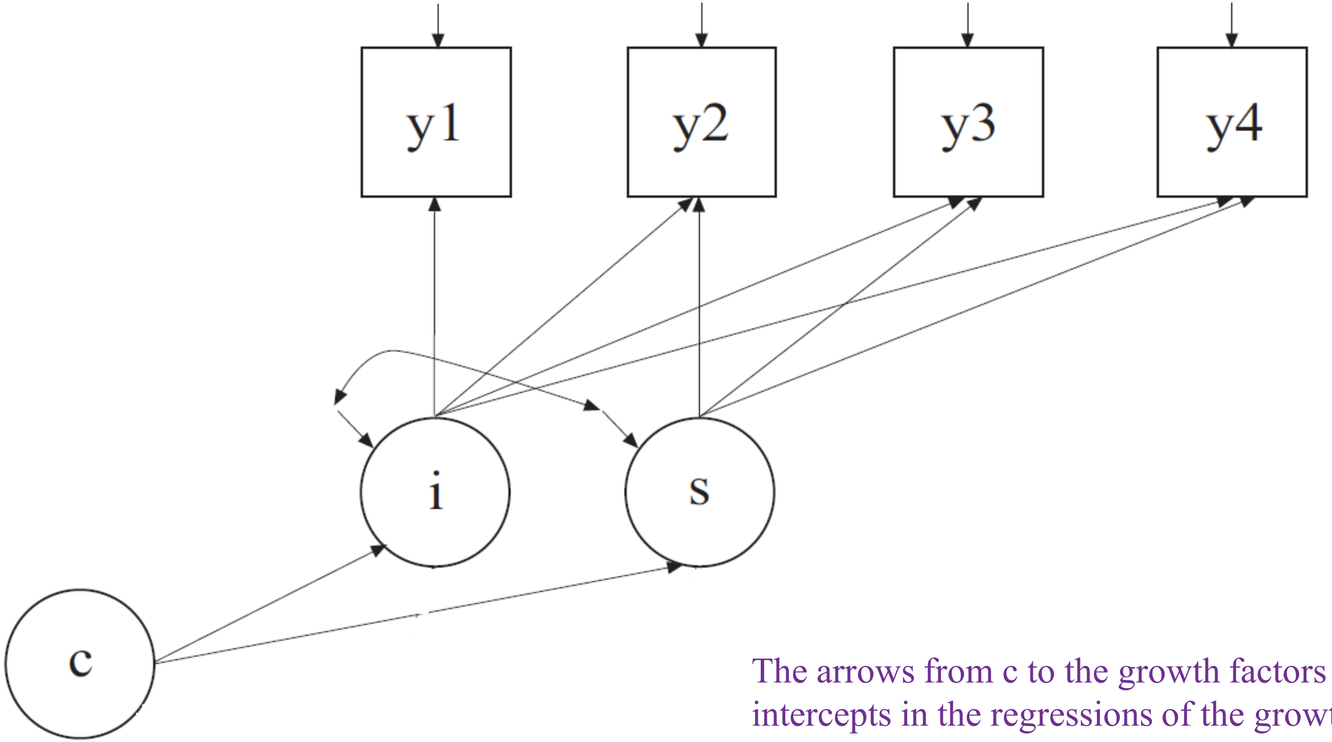
Multilevel (growth curve) model

```
TITLE:      this is an example of a linear growth
            model for a continuous outcome
DATA:       FILE IS ex6.1.dat;
VARIABLE:   NAMES ARE y11-y14 x1 x2 x31-x34;
            USEVARIABLES ARE y11-y14;
MODEL:      i s | y11@0 y12@1 y13@2 y14@3;
```



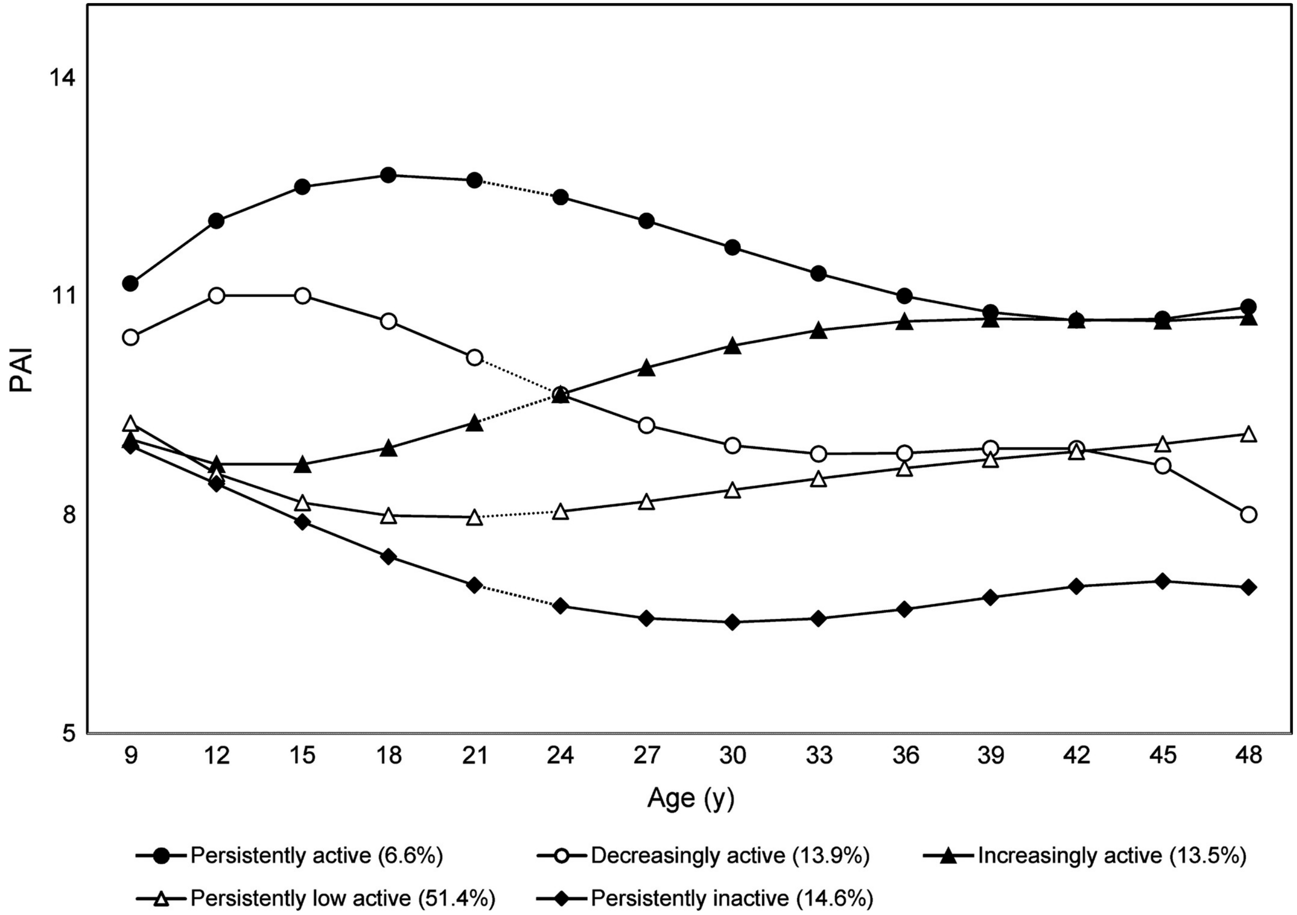
Latent (growth curve) model

Growth mixture model

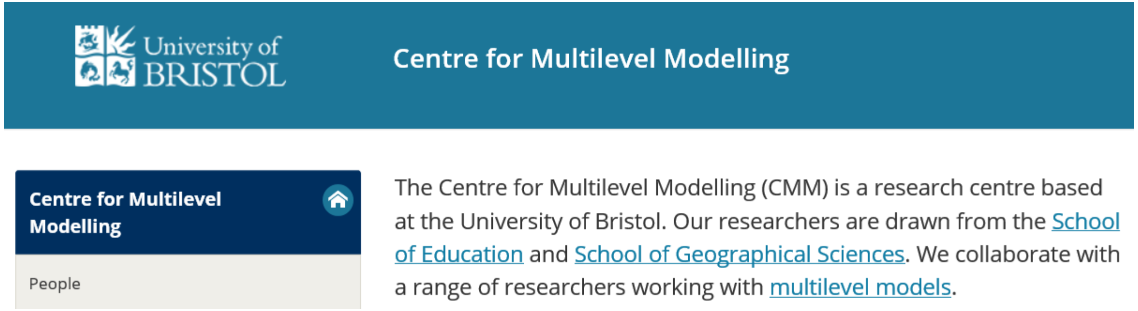


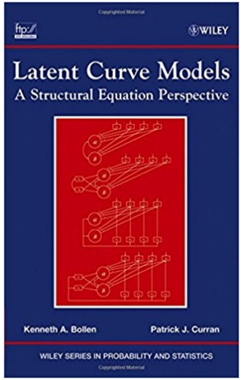
The arrows from c to the growth factors i and s indicate that the intercepts in the regressions of the growth factors on x vary across the classes of c. This corresponds to the regressions of i and s on a set of dummy variables representing the categories of c.

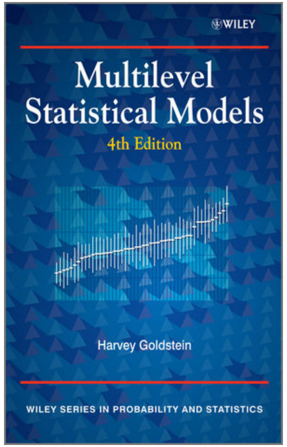
Growth mixture model



Resources







Johnson W. 2015. Analytical strategies in human growth research. Am J Hum Biol 27(1): 69-83

Tu YK, Tilling K, Sterne JA, Gilthorpe MS. 2013. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. Int J Epidemiol 42:1327-1339