A 3-step multilevel SEM

Yajing Zhu [1,2] Fiona Steele [2] Irini Moustaki [2]

[1]MRC Biostatistics Unit, University of Cambridge
[2]Department of Statistics, London School of Economics
7 June 2019

Substantive research question

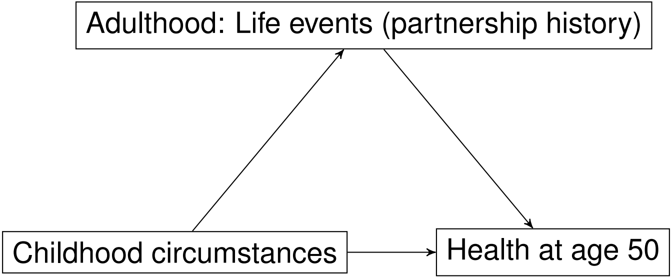


Figure 1: A general joint modelling framework to explore the potential pathways between childhood circumstances, partnership history and health in mid-life.

Motivation

Substantive literature (Galobardes et al., 2006; Ben-Shlomo et al., 2016; Cohen et al., 2010; Lacey et al., 2014)

- Sources of social and health inequalities in midlife: socioeconomic circumstances in early life
- Life course influences transmit via physical, behavioural and psychosocial pathways (conceptual frameworks/mechanisms)
- Conceptual foundation of statistical models.

Description of the dataset: recently published sweep NCDS9 (2013-2014, age 55) achieved 9,125 CMs

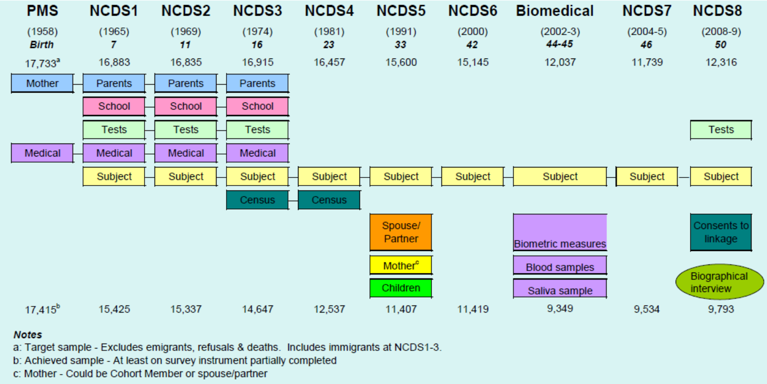


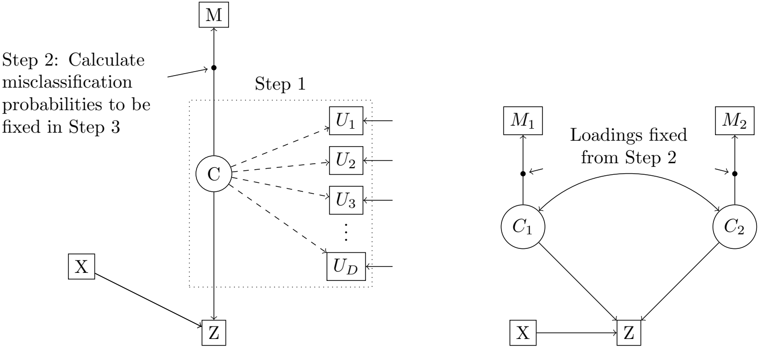
Figure 2: Overview of the dataset

Methodological challenges

- (Intermittent missing data, measurement error, multiple constructs) Childhood socio-economic circumstances (SECs) at ages 0, 7, 11 and 16
  - LCA $\rightarrow$ characterise the patterns of change in each dimension of childhood SECs
- Relate multiple categorical latent variables (LV) to temporally distal outcomes (clustered, mixed-type) $\rightarrow$ A general 3-step ML approach for multiple LVs (Zhu et al., 2017) [1]
  - To a single distal outcome
  - To time-to-event outcomes
  - In the multilevel SEM

[1] Note:Vermunt (2010); Asparouhov and Muthén (2014); Bakk and Vermunt (2016) contributed to the method for 1 LV.

3-step ML approach for SEM (1)



(a) One C - one distal outcome Z

(b) C_1 and C_2 - one distal outcome Z

Figure 3: Distal outcomes in mixture models

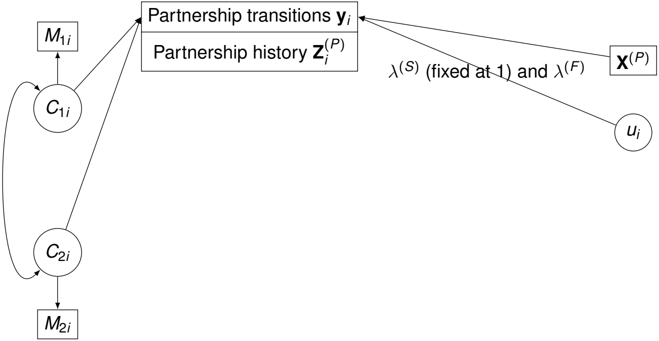


Figure 4: A general path diagram of a multilevel SEM for partnership transitions $\mathbf{y}_i = \{\mathbf{y}_i^{(F)}, \mathbf{y}_i^{(S)}\}$, distal health outcome H_i and the dropout mechanism $\mathbf{D}_i$ with factorised individual-level random effects u_i.

3-step ML approach for SEM (2)

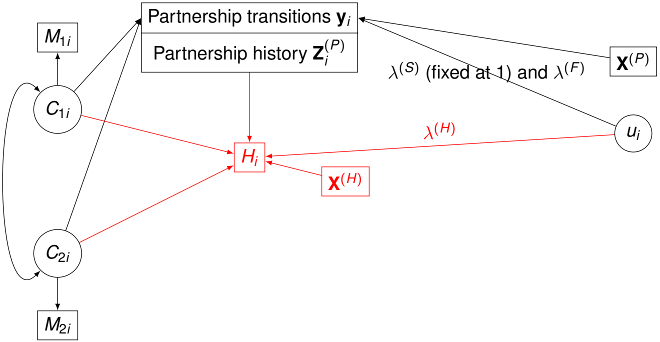


Figure 4: A general path diagram of a multilevel SEM for partnership transitions $\mathbf{y}_i = \{\mathbf{y}_i^{(F)}, \mathbf{y}_i^{(S)}\}$, distal health outcome H_i and the dropout mechanism $\mathbf{D}_i$ with factorised individual-level random effects u_i.

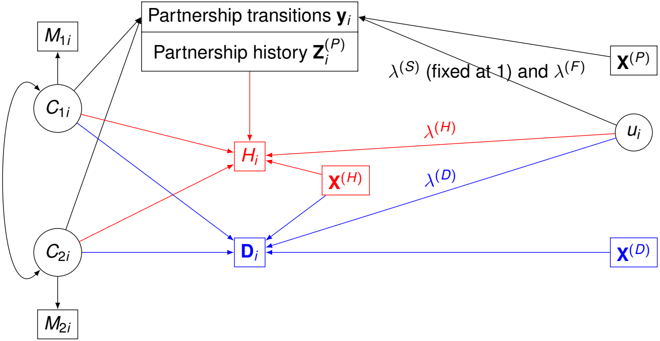


Figure 4: A general path diagram of a multilevel SEM for partnership transitions $\mathbf{y}_i = \{\mathbf{y}_i^{(F)}, \mathbf{y}_i^{(S)}\}$, distal health outcome H_i and the dropout mechanism $\mathbf{D}_i$ with factorised individual-level random effects u_i.

Discrete-time survival data (partnership transitions)

- Denote by y_{ij} the duration of episode j of individual i, which is fully observed if an event occurs ($\delta_{ij} = 1$) and right-censored if not ($\delta_{ij} = 0$).
- Data restructuring: convert the observed data (y_{ij}, δ_{ij}) to a sequence of binary responses (y_{tij}), indicating whether an event has occurred in time interval $[t, t + 1)$.
- Discrete-time hazard function: $h_{tij} = Pr(y_{tij} = 1 | y_{t'<t,ij} = 0)$.

Childhood SECs $\rightarrow$ time-to-event outcomes (2)

Step 3 is a random effects logit model, allowing for a log-linear structure between LVs.

$$\log\left(\frac{h_{tij}}{1-h_{tij}}\right) = \alpha_t + \boldsymbol{\beta}'\mathbf{X}_{tij} + \sum_{q=1}^{Q}\sum_{k_q=1}^{K_q-1} \tau_{C_q,k_q} I(C_{qi} = k_q) + u_i$$

- h_{tij} is the hazard of partnership transitions (formation and dissolutions)
- α_t is the baseline hazard function
- $\mathbf{X}_{tij}$ is the vector of time-varying and time-invariant predictors
- τ_{C_q,k_q}s are the class-specific coefficients of LV C_q
- $u_i \sim N(0, \sigma_u^2)$ is the individual-specific unobserved random effect

Multilevel SEM

Model specification

$$\text{logit}\left(h_{tij}^{(P)}\right) = \alpha_t + \sum_{k_q=1}^{K_q-1} \alpha_{C_q,k_q} I(C_{qi} = k_q) + \boldsymbol{\alpha}'\mathbf{X}_{tij}^{(P)} + u_i,$$

$$\text{logit}\left(p_i^{(H)}\right) = \beta_0 + \sum_{k_q=1}^{K_q-1} \beta_{C_q,k_q} I(C_{qi} = k_q) + \boldsymbol{\beta}_1'\mathbf{X}_i^{(H)} + \boldsymbol{\beta}_2'\mathbf{Z}_i^{(P)} + \lambda^{(H)} u_i.$$

$$\text{logit}\left(h_{ri}^{(D)}\right) = \alpha_r^{(D)} + \boldsymbol{\beta}^{(D)'}\mathbf{X}_i^{(D)} + \sum_{q=1}^{4} \sum_{k_q=1}^{K_q-1} \tau_{C_q,k_q}^{(D)} I(C_{qi} = k_q) + \lambda^{(D)} u_i.$$

- $p_i^{(H)} = p(H_i = 1)$, H_i is binary health status (1 for poor).
- $h_{ri}^{(D)} = P(D_{ri} = 1 | D_{r'<r,i} = 0)$, D_{ri} is a dropout indicator in each wave r.
- $\mathbf{X}_i^{(H)}$ is a vector of health-relevant covariates.
- $\mathbf{X}_{tij}^{(P)}$ is a vector of predictors of separation hazard.
- $\mathbf{Z}_i^{(P)}$ is a vector of summary indicators of partnership stability derived from the partnership history (e.g. $\#$ partners during ages 16-50, % time single).

Advantages of the framework

- Corrects for misclassification error
- Endogenous$\mathbf{Z}_i^{(P)}$ in the health model
- Differential random effects (λ)
- Conditional dependence between outcomes
- Generalisability: data with complex structures (e.g. multilevel, longitudinal, mixed response types), dropout mechanism and related processes, multiple health outcomes $\Rightarrow$ better identification of σ_u^2.

A 3-step multilevel SEM with a submodel for the time to dropout.

Heath submodel

Covariates	Est.	(SE)
Intercept	-3.06**	(0.23)
Overweight[1] (ref.= No)	0.26**	(0.07)
Childhood circumstances		
Social class[2] (ref.=High)		
Low	0.44**	(0.12)
Medium	0.30**	(0.10)
Financial difficulty (ref.=Low)		
High	0.42**	(0.10)
Material hardship (ref.=Low)		
Medium	0.32**	(0.09)
High	0.39**	(0.10)
Family structure (ref.=Stable)		
Unstable	0.17	(0.17)
Partnership experience		
Total number of partners before age 50 (ref. =1)		
0	-0.13	(0.32)
2	0.18	(0.14)
3+	0.41*	(0.24)
Age at first partnership	-0.13**	(0.05)
Percentage time spent single	1.26**	(0.38)
Random effect parameters		
σ_u^2	1.32**	(0.10)
$\lambda^{(H)}$	-0.44**	(0.16)
$\lambda^{(F)}$	-0.05**	(0.02)
$\lambda^{(D)}$	1.25**	(0.12)

** $p < 0.05$, * $p < 0.1$

[1] Binary indicator for overweight at age 16.

[2] Father or male head social class.

Effects of childhood SECs on partnership transitions and dropout

Covariates	First partnership Est.	First partnership (SE)	Dissolutions Est.	Dissolutions (SE)	Dropout Est.	Dropout (SE)
Social class[a] (ref.=High)						
Low	0.11**	(0.04)	-0.07	(0.07)	0.18**	(0.07)
Medium	0.11**	(0.03)	0.01	(0.05)	0.13**	(0.06)
Financial difficulty (ref.=Low)						
High	0.03	(0.04)	0.17**	(0.07)	0.30**	(0.07)
Material hardship (ref.=Low)						
Medium	0.04	(0.03)	-0.09	(0.06)	0.05	(0.06)
High	0.05	(0.03)	-0.11*	(0.06)	0.20**	(0.06)
Family structure (ref.=Stable)						
Unstable	0.15**	(0.03)	0.29**	(0.07)	0.30**	(0.08)

**$p < 0.05$, *$p < 0.1$.

[1] Father or male head social class.

Future work (medical sciences)

- Precision medicine: classification - risk prediction models
- Electronic health records (GP data, informative presence)
- Joint modelling of time-to-event data (multiple event histories), longitudinal data (multiple biomarkers), informative presence.
  - Rizopoulos (2011),Tsiatis and Davidian (2004), Crowther et al. (2016)
  - Dynamic risk prediction, landmarking.

Asparouhov, T. and Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3):329–341.

Bakk, Z. and Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1):20–31.

Ben-Shlomo, Y., Cooper, R., and Kuh, D. (2016). The last two decades of life course epidemiology, and its relevance for research on ageing. *International Journal of Epidemiology*, 45(4):973–988.

Cohen, S., Janicki-Deverts, D., Chen, E., and Matthews, K. A. (2010). Childhood socioeconomic status and adult health. *Annals of the New York Academy of Sciences*, 1186(1):37–55.

Crowther, M. J., Andersson, T. M., Lambert, P. C., Abrams, K. R., and Humphreys, K. (2016). Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine*, 35(7):1193–1209.

Galobardes, B., Smith, G. D., and Lynch, J. W. (2006). Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Annals of Epidemiology*, 16(2):91–104.

References (2)

Lacey, R. E., Bartley, M., Pikhart, H., Stafford, M., and Cable, N. (2014). Parental separation and adult psychological distress: an investigation of material and relational mechanisms. *BMC Public Health*, 14(1):272.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4):450–469.

Zhu, Y., Steele, F., and Moustaki, I. (2017). A general 3-step maximum likelihood approach to estimate the effects of multiple latent categorical variables on a distal outcome. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(5):643–656.