Understanding the life course impact of childhood socioeconomic disadvantages on midlife health: a multilevel SEM approach

Yajing Zhu
Prof. Fiona Steele, Prof. Irini Moustaki

MRC Biostatistics Unit
University of Cambridge, UK
Department of Statistics
London School of Economics and Political Science, UK

14 Jan. 2019

Outline

- My research interests
- Latent variable modelling for life course research
- Discussions

Research interests

- Methodology: latent variable models, longitudinal data analysis (multilevel models), event history models (survival models), measurement error, missing data, model misspecification (robustness), selection (endogeneity)
- Applications: life course/behavioural epidemiology, public health, sociology, demography, ageing, ++
- Data: national cohort studies (e.g. 1958 NCDS), longitudinal survey data (observational studies)
- Welcome collaborations/general discussions. Questions? Data?

Data I

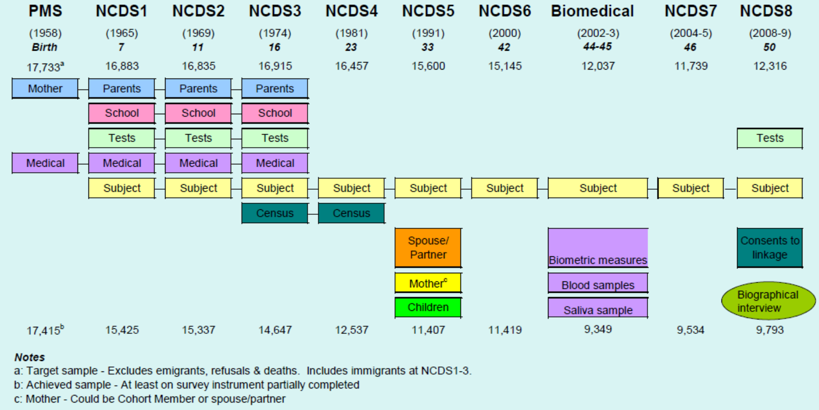


Figure 1: 1958 National Child Development Study: N=18,558, followed from birth to age 55.

Data II

- At ages 0, 7, 11 and 16: multiple and repeated measures of childhood circumstances (e.g. parental social class, housing and financial situations, parental relationship), physical health (e.g. height, weight, records of hospital episodes), other adverse experiences (e.g. neglect, abuse)
- Partnership, employment, fertility histories for ages 16-55 (recorded to months): event history data
- Multiple measures of health (self-reported, medical wave at age 45-46) in adulthood (ages 16-55, but only at interview times)

Data III

Main interest

- Bring together rich information on multiple concepts, at different life stages
  - Common approaches: include RM as predictors;ad-hoc composite summary; dimension reduction using multiple related indicators at different life stages
  - Problems: measurement error, inflated SE, overly broad summary (often in 1 LV), not capturing patterns of change over time.
- Relate time-to-event data (comprehensive event histories) in adulthood to distal outcomes (e.g. later health); to changes in health states over time
- Simultaneously explore dynamic relationships between quantities + cumulative effects of, e.g. life events on later health; early childhood circumstances on life experiences and on well-being in later life

Simplified framework: an example

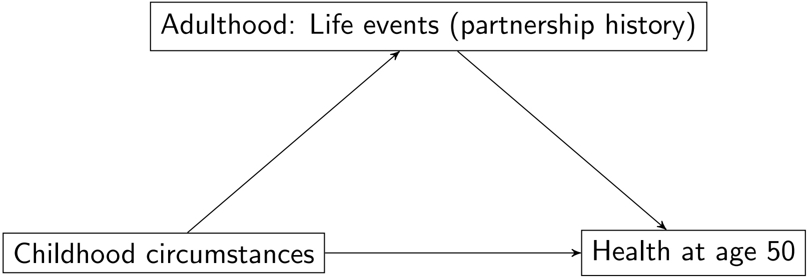


Figure 2: A general joint modelling framework to explore the potential pathways between childhood circumstances, partnership history and health in mid-life.

Motivation

- Substantive (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 Bartley (2016)
  - Life course influences transmit via physical, behavioural and psychosocial pathways (conceptual frameworks/mechanisms)
  - Conceptual foundation of statistical models
- Methodological
  - Large cohort data: 1958 British NCDS 10 waves for age 0-50 (hierarchical, mixed-type, longitudinal and event history data)
  - Measurement error, classification error, complex association structures, endogeneity, missing data
  - Multilevel SEM: interpretable, flexible

Challenges: the scope of the statistical analysis

- How should we measure childhood circumstances (multiple dimensions, abstract)?
- How do childhood circumstances influence partnership transitions between ages 16 and 50?
- How do partnership transitions mediate the effects of childhood circumstances on midlife health?
- Missing data (intermittent missingess, drop-outs)

Overview of the methodology I

1. (Intermittent missing data, measurement error) Latent class models to characterise the overall pattern of each of the 4 dimensions of childhood SECs (father's social class, financial difficulty, material hardship, family structure)
    - Use/construct repeated measures of each dimension at four childhood waves
    - Run LCA on each set of measures for each dimension $->$ each latent class variable reflects the pattern of change over childhood
    - Ready to relate these latent class variables (LV) to outcomes of interest

Overview of the methodology II

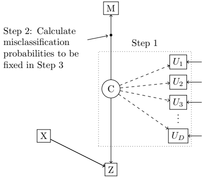
2. Relate LVs to temporally distal outcomes (at individual-level or time-domain, of mixed types) $\rightarrow$ A general 3-step ML approach for multiple LVs (Zhu et al., 2017) [1]

Attention

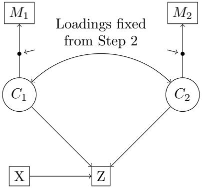
- Class assignments (e.g. most likely class membership) are not TRUE classes (i.e. imperfect measurements of true classes)
- Be aware of incurred statistical assumptions (distributional + functional + conditional independence)
- Too many dimensions? (Is there substantive necessity? Got a lot of data?)
- Measurement models (LCAs) need to be carefully examined for good fit

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

More on the general 3-step ML approach for SEM I



(a) 1-LV 3-step



(b) 2-LV 3-step

Figure 3: The 3-step approach where one or two (possibly associated) latent categorical variables C_1 and C_2 are related to a distal outcome Z.

More on the general 3-step ML approach for SEM II

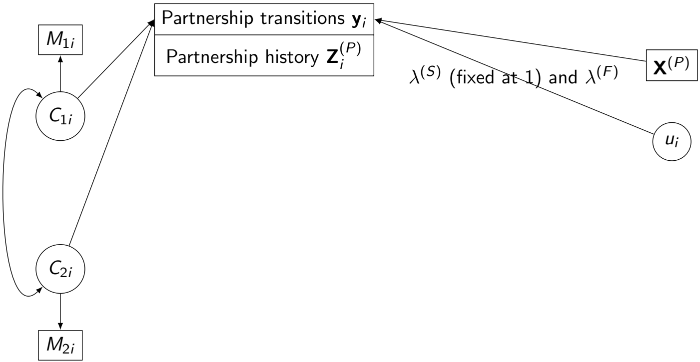


Figure 4: A general path diagram of a multilevel SEM for partnership transitions $\mathbf{y}_i = \{\mathbf{y}_i^{(F)}, \mathbf{y}_i^{(S)}\}$, distal health outcome H_i and the dropout mechanism $\mathbf{D}_i$ with factorised individual-level random effects u_i.

More on the general 3-step ML approach for SEM II

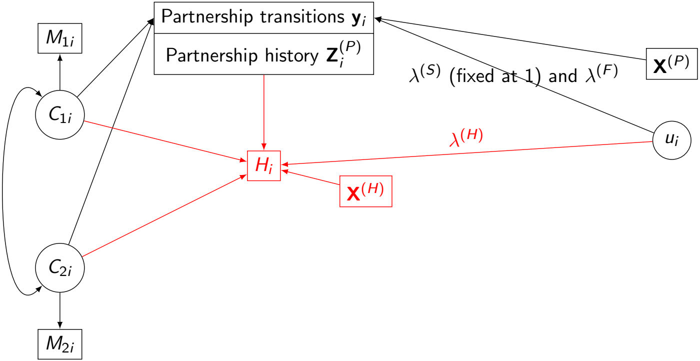


Figure 4: A general path diagram of a multilevel SEM for partnership transitions $\mathbf{y}_i = \{\mathbf{y}_i^{(F)}, \mathbf{y}_i^{(S)}\}$, distal health outcome H_i and the dropout mechanism $\mathbf{D}_i$ with factorised individual-level random effects u_i.

More on the general 3-step ML approach for SEM II

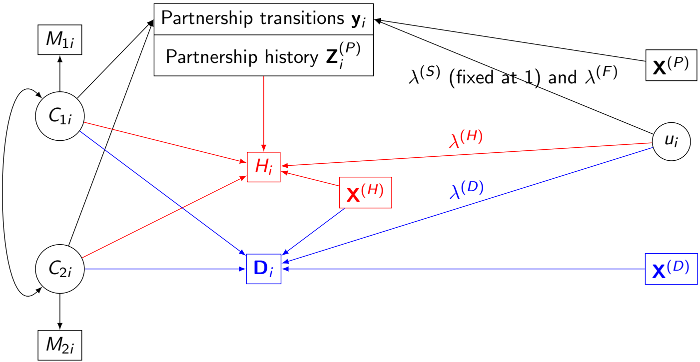


Figure 4: A general path diagram of a multilevel SEM for partnership transitions $\mathbf{y}_i = \{\mathbf{y}_i^{(F)}, \mathbf{y}_i^{(S)}\}$, distal health outcome H_i and the dropout mechanism $\mathbf{D}_i$ with factorised individual-level random effects u_i.

Practical issues: Estimation

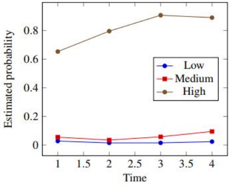
- Simulation study: R+LatentGOLD (batch mode); Empirical study: Mplus (manual approach for multiple LVs, automatic approach for 1 LV), LatentGOLD (automatic function)
- Empirical study: account for potential violation of model assumptions + a proper sensitivity analysis (misspecification: neglected important common predictors?)
- Too many LVs may overcomplicate the model (unless N large; large within-individual variability)

Substantive contributions I

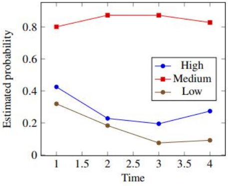
Key findings

- Childhood SECs: captured by four latent class variables, each derived by fitting LCA to the respective set of four repeated measures of one dimension of childhood SECs.
- Examples of estimated probabilities of each grouped categories of each childhood measure, conditional on class membership

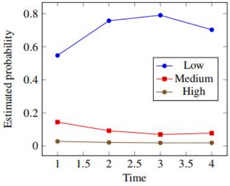
Substantive contributions II



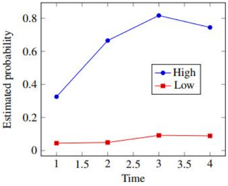
(a) Social class: Managerial/Professional



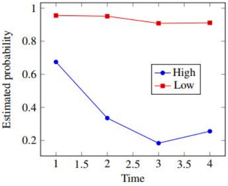
(b) Social class: Skilled



(c) Social class: Partially skilled/Unskilled/Unemployed



(d) Financial difficulty: Yes



(e) Financial difficulty: No

Substantive contributions III

- Material disadvantages (i.e. low level of social class, higher levels of financial difficulty and material hardships) in childhood have a significantly direct and persistent influence on a higher tendency of poor midlife health.
- Material disadvantages also have an indirect influence on midlife health by pushing up the risk of forming the first partnership and that of dissolutions.
- The impact of unstable family structure on midlife health is only indirect, through influencing partnership transitions.
- Early union formation and unstable partnership experience significantly increase the probability of poor midlife health.
- Residual interrelationships between partnership experiences, midlife health and dropout tendency are found. Non-ignorable dropout is confirmed.

Next steps

- Multiple distal health outcomes (possibly of mixed types)
- Multiple event histories
- Time-varying effects of childhood SECs on experiences of life events over time
- Impact of childhood SECs on the development of health trajectory
- Dynamic relationship between event histories and health (ages 16-50) - lagged models / LTA
- Joint modelling of time-to-event data (event histories) and longitudinal data (repeated measures of health)
    - Rizopoulos (2011),Tsiatis and Davidian (2004), Crowther et al. (2016)
    - Effects of $H_{t-1,i}$ (or the trajectory $\{H_{t'i} : t' < t\}$) on time-to-event outcomes (from t onwards)
    - Event occurrence at t may influence later health at t' ($t' > t$) (temporary/persistent)

Discussions

- Model fit: Not so straightforward. Global tests inappropriate (and do not tell you much); Prediction error (Not a good criteria if the main interest is to understand association/mechanism) $->$ local model fit and relative fit (compare nested models)
- Our approach: flexible to handle complex longitudinal data and model structures. Alternative (big field): causal inferences (Bianca) with DAG, potential outcome framework (challenges in the longitudinal setting: time-varying confounding, mediators (or mediating process), intermediate confounders)
- Impact. Many statistical problems to investigate, but which are of substantive impact to the society/to bring in new insights $->$ close collaboration with researchers with substantive insights!
- Robustness, generalisability, complexity vs. accuracy vs. interpretability

Statistical models

Denote by $P_{si}^{(F)} = P(y_{si}^{(F)} = 1 | y_{s'<s,i}^{(F)} = 0)$ (formation hazard), $P_{tij}^{(S)} = P(y_{tij}^{(S)} = 1 | y_{t'<t,i}^{(S)} = 0)$ (separation hazard), $P_{i}^{(H)} = P(y_{i}^{(H)} = 1)$ (probability of poor midlife health) and $P_{ri}^{(D)} = P(D_{ri} = 1 | D_{r'<r,i} = 0)$ (dropout hazard), and αs the time or wave-specific logit baseline hazard.

$$\begin{cases} \operatorname{logit}\left(h_{si}^{(F)}\right) = \alpha_s^{(F)} + \boldsymbol{\beta}^{(F)'}\mathbf{X}_{si}^{(F)} + \sum_{q=1}^{4}\sum_{k_q=1}^{K_q-1}\tau_{C_q,k_q}^{(F)} I(C_{qi} = k_q) + \lambda^{(F)} u_i \\ \operatorname{logit}\left(h_{tij}^{(S)}\right) = \alpha_t^{(S)} + \boldsymbol{\beta}^{(S)'}\mathbf{X}_{tij}^{(S)} + \sum_{q=1}^{4}\sum_{k_q=1}^{K_q-1}\tau_{k_q}^{(S)} I(C_{qi} = k_q) + u_i \\ \operatorname{logit}\left(P_{i}^{(H)}\right) = \alpha_0^{(H)} + \boldsymbol{\beta}_1^{(H)'}\mathbf{Z}_{i}^{(P)} + \boldsymbol{\beta}_2^{(H)'}\mathbf{X}_{i}^{(H)} + \sum_{q=1}^{4}\sum_{k_q=1}^{K_q-1}\tau_{C_q,k_q}^{(H)} I(C_{qi} = k_q) + \lambda^{(H)} u_i \\ \operatorname{logit}\left(P_{ri}^{(D)}\right) = \alpha_r^{(D)} + \boldsymbol{\beta}^{(D)'}\mathbf{X}_{i}^{(D)} + \sum_{q=1}^{4}\sum_{k_q=1}^{K_q-1}\tau_{C_q,k_q}^{(D)} I(C_{qi} = k_q) + \lambda^{(D)} u_i \end{cases}$$

References I

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.

Bartley, M. (2016). *Health inequality: an introduction to concepts, theories and methods*. John Wiley & Sons.

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 II

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.