Invited sessions for ISCB42 (Lyon, 2021)
Session 2

Variance modelling for multilevel data and joint models

Organizers : Donald Hedeker : University of Chicago, US & Karen Leffondré (Bordeaux University, FR)

Michael Crowther : Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm
A general framework and implementation for variance modelling in joint model settings

Ellen Hamaker : Department of Methodology and Statistics, Utrecht University
Dynamic structural equation modeling for intensive longitudinal data

Jessica Barrett : Medical Research Council Biostatistics Unit, University of Cambridge
Jointly modelling longitudinal heteroscedasticity and a time-to-event outcome

A general framework and implementation for variance modelling in joint model settings

Michael J. Crowther[1]

[1] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

The rise in availability of electronic health record data enables us to answer more detailed clinical questions; however, the associated increased complexity raises substantial statistical and computational challenges. Recent work in the area of joint models has introduced an extended mixed effects framework, encompassing multiple outcomes of any type, each of which could be repeatedly measured (longitudinal), with any number of levels, and with any number of random effects at each level (Crowther, 2020). This allows for sharing and linking between outcome models in an extremely flexibly way, either by linking random effects directly, or the expected value of one outcome (or function of it) within the linear predictor of another. Non-linear and time-dependent effects are also seamlessly incorporated to the linear predictor through the use of splines or fractional polynomials. In this talk, I'll present an extension to the framework to further allow modelling of variance components directly, allowing each random effect or residual variance to have its own complex linear predictor, such as allowing for heteroskedasticity, which in turn provides new tools for joint modelling. Throughout my talk I will illustrate an accompanying user-friendly implementation in Stata, showing how to build and estimate a joint longitudinal-survival model with complex variance components, quantifying how between-subject variation in the level 1 variance structure of a continuous biomarker (e.g., blood pressure), can be associated with survival. Dynamic predictions from such a joint model will also be derived and presented. Due to the generality of the implementation, multiple outcomes, such as multiple biomarkers or competing risks, are also immediately available.

Crowther MJ. merlin - a unified framework for data analysis and methods development in Stata. *Stata Journal* 2020;20(4):763-784.

Dynamic structural equation modeling for intensive longitudinal data

Ellen L. Hamaker[1].

[1] Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands (presenting author)

Objective: Recent methodological innovations have opened up an exciting new horizon of research opportunities. Technological developments such as smartphones and other wearable devices have created new data collection methods such as ambulatory assessments, experience sampling, and ecological momentary assessments. Characteristic of the intensive longitudinal data that are obtained with these methods is that they consist of large numbers of repeated measurements of what individuals are doing, thinking and feeling while living their daily life. As such, these data provide us the opportunity to study individual processes as they unfold over time, and to investigate individual differences therein. But to fully realize this unique potential of intensive longitudinal data, we need new statistical techniques that adequately deal with the specifics of these data, and that can uncover the meaningful patterns hidden in them. This has led to diverse innovations, including the development of dynamic structural equation modeling (DSEM), a new toolbox in the software package Mplus.

Statistical Methods: DSEM forms a combination of: a) *time series analysis* to model the lagged relations between variables, thereby accounting for the autocorrelation structure within individuals; b) *multilevel modeling* to ensure the proper decomposition of variance into within-person and between-person components, and to allow for individual differences in means, slopes, and variances; and c) *structural equation modeling* such that multiple observed variables can be analyzed simultaneously, and can be combined using factor analysis and/or mediation analysis. Additionally, DSEM can account for unequal time-intervals between observations, and for trends and cycles over time.

Application: In this talk, I will briefly sketch the general DSEM framework, and then discuss several specific ways in which DSEM can be used to analyze particular data structures and tackle specific research questions. These include data from a design in which a randomized controlled trial is combined with intensive longitudinal measurements before and after the intervention, and data for which there may be a need to account for patters due to cycles such as day-of-the-week patterns or circadian rhythms. These empirical examples illustrate the wide variety of modeling opportunities that are offered by DSEM.

Jointly modelling longitudinal heteroscedasticity and a time-to-event outcome

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

[1] MRC Biostatistics Unit, University of Cambridge, UK (presenting author)
[2] MRC Integrative Epidemiology Unit, University of Bristol, UK
[3] ENSAE, France
[4]Department of Preventive Medicine, University of Mississippi Medical Center, USA
[5]Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, USA

In the clinical literature it has been shown that individuals with higher variability in their blood pressure measurements have a greater risk of cardiovascular disease. This is typically explored by calculating a variability measure, e.g. the standard deviation, from a set of blood pressure measurements per individual, and including this as an explanatory variable in a regression model for the time to the first cardiovascular event. However, this leads to regression dilution bias in the estimated association parameter because the variability measure is subject to measurement error.

We will explore statistical models which allow within-individual variability, as well as the mean, to depend on covariates and/or random effects, e.g. mixed effects location scale models (Hedeker et al, 2008). We propose a joint model with mixed effects location scale and time-to-event sub-models for the longitudinal blood pressure measurements and time to first cardiovascular event respectively (Barrett et al, 2019). The time-to-event sub-model incorporates the random effect associated with the longitudinal within-individual variability, which allows direct estimation of the association between blood pressure variability and the risk of CVD events.

We use simulation studies and data from the Atherosclerosis Risk in Communities (ARIC) study to compare the joint model with the usual method used in the literature and a two-stage method. We demonstrate that substantial bias may be incurred by the usual method and slight to moderate bias with the two-stage method, especially when blood pressure measurements are taken concurrently with the time-to-event follow-up. From the analysis of ARIC study data, the estimated hazard ratio for the association between visit-to-visit systolic blood pressure variability and cardiovascular disease from a joint model with random intercept, slope and variability effects is 1.08 (95% CI 1.04, 1.09) per unit increase in systolic blood pressure standard deviation.

References

1. Hedeker et al (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics* **64**: 627-634.
2. Barrett et al (2019). Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study. *Statistics in Medicine* **38**:1855-1868.