Universiteit Utrecht

# 12th International Multilevel Conference
# April 9 & 10, 2019

## Conference Program
## & Abstracts


**Organizing committee**
Mirjam Moerbeek
Rens van de Schoot
Leoniek Wijngaards – de Meij
Marianne Geelhoed *(local organization)*

Utrecht University
Department Methodology & Statistics

# Conference program

## Day 1 (Tuesday April 9) - morning

| | |
|---|---|
| | **Room: Kerkzaal** |
| 09:00 | **Registration** |
| 09:25 | **Opening** |
| 09:30 | **Keynote 1: Todd Little**<br>On the Merits of Longitudinal Multiple-Group Fixed Effects Modeling Versus Multilevel Modeling for Evaluating Interventions. |
| 10:00 | *Fayette Klaassen*<br>(Not) Everybody Does: Testing for Individual Differences and Similarities in Hierarchical data. |
| 10:20 | *Yongyun Shin*<br>Income Equality in Achievement among US Elementary Schools: A Random Coefficients Model with Data MAR. |
| 10:40 | **Coffee and Tea Break** |
| 11:00 | *Claire Durand*<br>How to Compare Data from Very Different Sources: A 4-level Longitudinal Model of Institutional Trust. |
| 11:20 | *Steven Teerenstra*<br>Sample size formulas for cluster randomized repeated measurement designs with p>2 levels. |
| 11:40 | *Jitske Sijbrandij*<br>Optimal developmental trajectory group analyses: Which parameters should (not) be constrained to accurately estimate growth mixture models? |
| 12:00 | *Richard Parker*<br>A joint modelling approach to relate within-individual variability in a repeatedly measured exposure to a future outcome, allowing for measurement error in the repeated measures. |

## Day 1 (Tuesday April 9) - afternoon

| | |
|---|---|
| | **Room: Kerkzaal** |
| 12:20 | **Lunch – Poster session** |
| 13:20 | *Alvaro Fuentes*<br>Multilevel Propensity Scores: An Evaluation of Findings. |
| 13:40 | *George Leckie*<br>Calculating intraclass correlation coefficients in multilevel models for count responses. |
| 14:00 | *Mariska Barendse*<br>On the use of pairwise maximum likelihood estimation for clustered data. |
| 14:20 | **Short break** |
| 14:30 | *Alexander Schmidt-Catran*<br>Why country dummies sometimes do not do the job. How to get the within-estimator of cross-level interactions with pooled cross-sections. |
| 14:50 | *Carla Rampichini*<br>Multiple imputation and selection of ordinal level-2 predictors in multilevel models: analysis of the relationship between student ratings and teacher beliefs and practices. |
| 15:10 | *David Wutchiett*<br>Missing data imputation in large combined cross-sectional and longitudinal data: multilevel multiple imputation and time series imputation. |
| 15:30 | **Coffee and Tea Break** |
| 15:50 | *Alice Richardson*<br>Multiple Imputation in Three-level Models. |
| 16:10 | *Simon Grund*<br>Multiple imputation of missing data in multilevel models with random slopes and nonlinear effects. |
| 16:30 | **Keynote 2: Stef van Buuren**<br>Recipes for multilevel imputation. |
| 17:00 | **End of day 1** |
| 17:00 | **Drinks and conference dinner *(for those who registered)*** |

## Day 2 (Wednesday April 10) - morning

| | |
|---|---|
| | **Room: Kerkzaal** |
| 09:00 | **Doors open** |
| 09:30 | **Keynote 3: Paul Bürkner**<br>Bayesian Multilevel Modeling with brms and Stan. |
| 10:00 | *Bill Browne*<br>Developing a statistical analysis assistant for Small Area Estimation in StatJR. |
| 10:20 | *John Hendrickx*<br>Using R to Evaluate Collinearity in Mixed Models. |
| 10:40 | **Coffee and Tea Break** |
| 11:00 | *Hawjeng Chiou*<br>Detecting Effects of Age, Period and Cohort on Growth Trajectory using Multilevel Modeling: Examination of Wage Trajectory of 1999-2016 in Taiwan. |
| 11:20 | *Marielle Zondervan-Zwijnenburg*<br>Testing replication of structural equation models. |
| 11:40 | *Justine Loncke*<br>The social relations model for count data: To Bayes or not to Bayes. |
| 12:00 | *Jean-Paul Fox*<br>Bayesian Covariance Structure Modelling. A novel method for multilevel data demonstrated with simulation and real data studies. |
| 12:20 | **Lunch** |
| | **Young Researcher Award nominees** |
| 13:20 | *Wouter Smink*<br>Assessing Individual Change Processes Bayesian Covariance Structure Modelling for negative associations among patients with personalized treatments. |
| 13:40 | *Tessa Johnson*<br>Modeling Student Mobility Using Hierarchical Networks. |
| 14:00 | *Fien Gistelinck*<br>Modeling longitudinal dyadic data in the SEM framework. |
| 14:20 | **Coffee and Tea Break** |
| 14:40 | *Yi Feng*<br>Variability as an Outcome Variable: Using Multilevel SEM to Model Lower-level Random Effect Variance<br>as Higher-level Latent Variables. |
| 15:00 | *Wendy Harrison*<br>Multilevel latent class (MLC) modelling of simulated upper-level causal effects in observational data. |
| 15:20 | *Sarah Chadwick*<br>Experimental design for multi-level data: Improving our approach to power analysis using Monte Carlo simulation-based parameter recovery estimation. |
| 15:40 | **PhD-award ceremony** |
| 16:00 | **End of day 2** |

# Abstracts *(in alphabetic order)*

# On the use of pairwise maximum likelihood estimation for clustered data

Barendse, M.T.[1*], Rosseel, Y. [1]

[1] Department of Data Analysis, Ghent University, Belgium

### Suggested talk duration (30 minutes)

### Summary

Social and behavioural research frequently involves multilevel data, with individuals and groups defined at separate levels. Multilevel analysis within the structural equation modeling framework often leads to the use of models with a large number of (latent) variables (i.e., random slopes, random intercepts, and hypothetical constructs) on different levels. The analysis of categorical multilevel data requires the evaluation of high-dimensional integrals. Current full-information approaches typically involve computationally intensive numerical methods (e.g., adaptive Gauss-Hermite quadrature or Markov chain Monte Carlo procedures). Alternatively, in the pairwise likelihood (PML) approach, the full likelihood is replaced by a sum of (bivariate) pairwise likelihoods, which are easier to handle. PML estimation has already been proven to be quite successful in single level datasets with a small number of categorical variables. In this presentation, we will show various possibilities of PML estimation for clustered data. Our approach is an extension of the so called 'wide' or 'multivariate' format approach that has been investigated by Bauer (2003), Curran (2003), and Mehta & Neale (2005) for continuous data with a multilevel structure.

### Relevance to conference theme

### Keywords (max. 3)

# Developing a statistical analysis assistant for Small Area Estimation in StatJR

Browne, W.J.[1*], Charlton, C.[1], Tzavidis, N.[2], Schmid, T.[3]

[1] University of Bristol, UK

[2] University of Southampton, UK

[3] Freie Universitat, Berlin, Germany

* Presenting author

**Suggested talk duration (20-30 minutes)**

**Summary (max. 500 words)**

Small Area Estimation (SAE) is a technique that is used in many application areas but often in official statistics. It is typically used when we have collected sample survey data from a population and we wish to produce estimates of a particular variable for groups, typically geographical areas, spanned by the population. Usually the survey is designed to only give population level estimates of the variable of interest and so SAE uses an additional census dataset on the whole population (not including the variable of interest) to impute data by using the relationship between the variable of interest and predictor variables that are present in both the survey and census data.

Tzavidis et al. (2018) give a framework that can be used to produce “classical” small area estimates using their emdi package (Kreutzmann et al., 2017) within R. We have been fortunate to obtain a collaborative grant from the ESRC to work on SAE using MCMC estimation and also to investigate interoperability with other software using our StatJR software (Charlton et al., 2013)

In this talk I will describe some of the work in the grant including: how we constructed an efficient MCMC algorithm using parallel processing; how StatJR can interoperate with R and the emdi package; and how through

creating a statistical analysis assistant we are able to embed the multilevel model fitting that occurs in unit-level SAE into a larger workflow and contains both pre and post estimation outputs useful for the applied researcher.

## Relevance to conference theme

Small Area Estimation for unit level models uses as the basis of the modelling multilevel modelling with some added imputation steps.

## Keywords (max. 3)

Small Area Estimation, StatJR, MCMC

# Experimental design for multi-level data: Improving our approach to power analysis using Monte Carlo simulation-based parameter recovery estimation

Chadwick, S.[1*],(PhD-student, primary supervisor: Davies, R) Davies, R.[1]

[1] Department of Psychology, Lancaster University, United Kingdom

**Suggested talk duration (15-60 minutes)**

15 minutes

**Summary (max. 500 words)**

Ensuring that our experimental studies are adequately powered to detect an effect of interest is a central concern across scientific disciplines. Multi-level experimental designs present a particular challenge for power analysis in its traditional sense, with most formulaic power analysis calculations silent to these internal data structures. In recent years, simulation-based approaches to power analysis have become more accessible, through R packages such as SIMR (Green & MacLeod, 2016). While this represents a welcome step-change, in their current form, these new approaches are limited in a number of ways. Power analysis is generally defined as the ability to recover an effect different from 0, involving a significance test in which a parameter estimate from a simulation-model is contrasted with null. In effect, this approach can tell us if our effect of interest is different from 0, but not by how much. Additionally, more user-friendly simulation-based power analysis methods typically offer limited flexibility in the range of model classes they can accommodate. I will demonstrate a general framework that can be used to overcome these issues: allowing for effective and informative calculation of parameter recovery across a range a model classes, including Bayesian approaches.

Reference:

Green, P. & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7(4)*, 493–498. DOI: https://doi.org/10.1111/2041-210X.12504

## Relevance to conference theme

Multi-level structures are hugely prevalent in experimental data. Being able to handle these structures in evaluating experimental design is a critical concern across scientific fields.

## Keywords (max. 3)

Power analysis, parameter recovery

# Detecting Effects of Age, Period and Cohort on Growth Trajectory using Multilevel Modeling: Examination of Wage Trajectory of 1999-2016 in Taiwan

Hsin-Lei Tseng[1] and Hawjeng Chiou[2*]

[1] Graduate Institute of Global Business and Strategy, National Taiwan Normal University, Taiwan

[2] College of Management, National Taiwan Normal University, Taiwan

## Summary

The estimation of age, period and cohort (APC) effects in a single model is always a challenge to longitudinal studies due to the restriction of the fully dependency of these variables. Such restriction causes identification problems in the traditional regression models. In terms of the advantage of decomposing effects into different levels in multilevel modeling, it allows age, period and cohort variables to be included with covariates at different levels. For predicting a growth curve such as wage trajectory, the period and cohort variables could be treated as a within-effect (level-one) and a between-effect (level-two) respectively, and age variable involves not only within- but between-effects in a traditional two-level model. This study demonstrated the application of multilevel modeling to identify age, period and cohort effects on wage trajectory of Taiwanese sample. The longitudinal data with 16 waves spanning 18 years of over 5,800 individuals in a Panel Study of Family Dynamics (PSFD) database was used. Premium effects of human capital factors, such as gender, educational levels and working hours, were also taken into account. Results showed that age was a significant predictor with a curvilinear trajectory with wage change which peaks at one's 50s across life span. Period was another effect with significant variations in one's wage trajectory which demonstrated its lowest point around the year of 2009 and defers from 2009 to 2014 when some covariates were controlled. However, although cohort effect revealed the highest cohort groups of earnings in 1966 to 1970, it became insignificant as age and period were simultaneously included. Moreover, levels of education also differentiated wage levels among individuals, while gender differences shown slightly influence on Taiwanese wage trajectory.  The methodological as well as managerial implications on the study of wage growth with APC effects were discussed in this study.

**Relevance to conference theme**

Application of multilevel modeling on the longitudinal data

**Keywords**

wage trajectory, age-period-cohort analysis, longitudinal data analysis

# How to Compare Data from Very Different Sources: A 4-level Longitudinal Model of Institutional Trust

Durand, C.[1], Peña Ibarra, L. P.[2]
[1] Université de Montréal, Canada
[2] Université de Montréal, Canada
* Claire Durand

## Suggested talk duration (20-30 minutes)

## Summary (max. 500 words)

Comparative research using data from the International Survey programs has often focussed on Western Europe and North America where it is rather easy to find similar measures of the concepts of interest. However, research is restricted by the fact that comparison is performed only on similar measures and in a context where there is not much variation between countries. In this paper, our aim is to compare trust in various institutions in regions where there is much variation between countries and where major changes in governance took place in the recent decades, i.e., in South and Central America, Sub-Saharan Africa, West Asia and North Africa, Asia and Eastern Europe.
In the selected regions, 17 International survey projects comprise measures of institutional trust. The main projects are the Barometers, the World and European Value Surveys, the European Social Survey, Life in transition and the Latin American Public Opinion Project (LAPOP). The measures of trust are disparate. The answer scales vary from 4-anchor scales to 5, 7, 10 and 11-point scales. More importantly, the institutions on which trust is assessed – more than 150 institutions and organizations -- vary between projects, between countries and over time. This is a typical case where multilevel analysis appears the only way to go. Answers to questions on trust for various institutions (level 1) are conceived as nested within respondents (level 2), themselves nested within surveys (the country-year level 3) and within countries (level 4). The data set combines more than 1370 surveys conducted in 143 countries since 1991. It comprises more than 1.8 million respondents and 22 million measures of trust.
This paper shows how the use of a 4-level multilevel longitudinal analysis of repeated measures allows for dealing *a posteriori* with harmonization issues. In order to deal with the variety of answer scales, we harmonize them on a unique scale and we control for the length of the original scale at the survey level. We can keep all the measures of trust in different institutions by nesting answers within respondents. Since there are too many institutions, the various institutions are grouped by theme a posteriori into 14 large categories of institutions in four different spheres, i.e. political, economical, administration and civil society. In the end, we can compare the level of trust between grouped institutions, between respondents, over time and between regions of the world. Cross-level interactions allow for testing whether trust in different institutions change similarly over time and in different regions. External indicators of the socio-political and economic situation of the different countries are merged with the data set in order to see whether they can explain some of the variation in trust between countries.
The paper will present the process used, its pitfalls and advantages, and the results of these analyses. It will conclude on the usefulness of this approach to study other research topics.

## Relevance to conference theme

The presentation fits in "innovative application" and, to a lesser degree in the "methodology" theme. It uses a 4-level longitudinal model of repeated measures, a type of model rarely if ever documented in the literature. It proposes a solution to difficult harmonization problems.

## Keywords (max. 3)

4-level multilevel analysis; longitudinal analysis; repeated measures

# Variability as an Outcome Variable: Using Multilevel SEM to Model Lower-level Random Effect Variance as Higher-level Latent Variables

Yi Feng[1*] & Gregory R. Hancock[1]

[1] University of Maryland, College Park, USA
* Presenting author (PhD student, supervisor/academic advisor: Dr. Gregory R. Hancock)

**Suggested talk duration (15-60 minutes)**

**Summary (max. 500 words)**

In many research and applied settings, it is variability, rather than means, that is of key interest. Examples include the stability of an individual's status over multiple measurements (e.g., across context or time), or cohesiveness/homogeneity across members within a team. More importantly, the goal may be to understand what factors have an impact on individual status stability or team cohesiveness, and/or how these variability-related characteristics would affect other distal outcomes. Such research questions naturally involve data that are multilevel in structure, and thus call for random effects models that partition the observed variance into variance components at different levels. Unlike traditional multilevel modeling scenarios, however, what is of focal interest here is the variability of the lower-level random effects variance across higher-level units—hence, a higher-level random effect.

Multilevel SEM (MSEM) has been shown to be a powerful tool, with performance comparable to traditional multilevel modeling but with the many versatility advantages of SEM (e.g., Bauer, 2003; Curran, 2003). With MSEM, the random effect can be conveniently modeled as a latent variable, and as such random effects can be embedded within a more general latent variable framework and incorporated in a broader causal modeling structure. Of direct relevance to the current project is recent work by Stapleton, Yang, and Hancock (2016), who proposed a way of modeling the cluster-specific within-cluster variance as a path coefficient through the introduction of a standardized phantom latent variable. This phantom path coefficient is, in turn, modeled as a latent factor at the cluster level with random effects. Overall, this capability of MSEM to model the variance of lower-level random effects as random latent variables at the higher level makes it a promising strategy to answer research questions with variability as the key outcome (or mediator).

In the current work we propose an analytical framework – one building upon multilevel modeling, SEM, and MSEM – whereby researchers can directly model the variability (as well as the mean) of the lower-level random effects variance as estimable parameters, and where the higher-level random effect of the lower-level random effects variance can be embedded in a more general latent variable modeling framework. We will first present the derivation of the proposed model, including both mathematical equations and graphical representations, followed by a discussion of estimation methods of key model parameters. Briefly, maximum likelihood (ML) estimation is directly applicable for two-level models that only involve manifest variables.

Based on cluster-specific variance-covariance matrices and mean vectors, we can obtain the ML estimates by minimizing an aggregated objective function (e.g., Rosseel, 2017). As the computation burden increases, such as when research questions explicitly involve measurement models, we propose Bayesian estimation methods (e.g., Asparouhov & Muthén, 2016; Stapleton et al., 2016). In addition to the theoretical development, we demonstrate the model fitting and parameter estimation process using several illustrative examples with empirical as well as simulated data. Further extensions will be discussed, such as conditional models, multiple group models, and longitudinal models for the change of intra-person stability or intra-cluster cohesiveness over time.

**Relevance to conference theme**

In this study, an innovative MSEM framework that draws upon both multilevel modeling and structural equation modeling is proposed. This proposed approach falls under the umbrella of multilevel modeling, while enjoying the many advantages of latent variable modeling framework. With this proposed analytic framework, multilevel modeling techniques can be applied to answer a wider range of research questions that have been outside the scope of the traditional modeling approaches. It is thus theoretically appealing, and provides applied researchers with a modeling tool that has a wide range of immediate implications.

**Keywords (max. 3)**

multilevel SEM, within-cluster cohesiveness, within-individual stability

References

Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163-192). Charlotte, NC: Information Age Publishing, Inc.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 135-167.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*, 529-569.

Rosseel, Y. (2017). Multilevel Structural Equation Modeling with lavaan [pdf slides]. Retrieved from http://users.ugent.be/~yrosseel/lavaan/zurich2017/MULTILEVEL/lavaan_multilevel_zurich2017.pdf

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41*, 481-520.

# Bayesian Covariance Structure Modelling

## A novel method for multilevel data demonstrated with simulation and real data studies

prof. dr. Jean-Paul Fox[1*]

Wouter A.C. Smink[1,2], MSc

[1] Department of Research Methodology, Measurement & Data Analysis

University of Twente

Enschede, The Netherlands

[2] Department of Psychology, Health & Technology

University of Twente

Enschede, The Netherlands

* Presenting and corresponding author, j.p.fox@utwente.nl

**Keywords**

- Bayesian Covariance Structure Modelling (BCSM)
- Small datasets
- Process data

### Suggested duration

- talk 30 minutes
- discussion 10 minutes

### Potential questions for discussants

- When are multilevel data considered to be small?
- Can the modeling framework be used to improve power and sample size computations in multilevel designs?
- Can the BCSM be used to develop small area predictions?
- Does the BCSM require the use of informative priors?

### Relevance to conference theme

As the *International Multilevel Conference* (IMC) covers statistical and methodological aspects of multilevel modelling, we feel that the IMC hosts an exceptional broad audience. We intend to interest both researchers and methodologists with our introduction in *Bayesian Covariance Structure Modelling* (BCSM) by giving a non-technical demonstration. We will elaborate on the estimation and interpretation of the BCSM through intuitive and realistic examples coming from a (simple) simulation study and real-data. Both our examples are aimed and enhancing the understanding of the advantages of BCSM they will be applicable to researchers coming from various scientific disciplines.

### Preferably in the same track

- Smink, Fox, Sools, Tjong Kim Sang, de Vries, Veldkamp, & Westerhof (2019). Assessing Individual Change Processes: Bayesian Covariance Structure Modelling for negative associations among patients with personalized treatments

## Summary

*The problem of small datasets*

Datasets in the social and medical sciences are often hierarchically structured: variables describe individuals, and individuals (often) adhere to groups. At the same time, data in these sciences remind us that not all data is *Big Data*: small samples are by no means a rare occurrence. Nevertheless, researchers often have comprehensive theories available, which lead into the direction of testing many parameters with multiple and complex dependencies.

Should small samples then be left aside by quantitative researchers? After all, a small dataset does not imply a lesser degree of importance, as there are a variety of reasons why datasets could be small. Correct statistical modelling is perhaps even more important when, for example, the population of the target group is extremely sparse (e.g., babies with a life-threatening orphan disease), or difficult to access (e.g., toddlers with autism from refugees).

Small datasets are especially challenging in multilevel modelling, as the sample size restrictions apply to each (modelled) hierarchical level in the data. Limited sample sizes greatly constrain meaningful statistical inference, as the sample determines the sufficient number of clusters (usually too few), and the size of the clusters themselves (usually too small). To overcome these issues, researchers often simplify their hypotheses and corresponding statistical models. Instead of doing that, we propose *Bayesian Covariance Structure Modelling* (BCSM), a novel method to deal with small data.

*Bayesian Covariance Structure Modelling for small datasets*

The main advantage of the BCSM (for small data) is that the covariance structure represents a random effect structure, but the random effects themselves do not have to be estimated. As the estimation of random effects is difficult with small data, this greatly improves the generalizability of conclusions.

A second advantage of the BCSM is that the model is reparametrized differently, avoiding the inclusion of random effects. Therefore, the number of parameters is drastically lower than the standard multilevel modeling approaches, while the interpretation does not change. Thus, BCSM allows for modelling of complex theories with limited data.

*Presentation*

We start with a basic simulation study, where we show the limitations of the standard multilevel modelling techniques when applied to small samples. We then showcase the advantages of the BCSM. This simulation will function as an useful introduction and reference point aimed at enhancing the understanding of the BCSM. We will then present the statistical issues and methodological advances with a basic BCSM using a real-data example.

# Multilevel Propensity Scores: An Evaluation of Findings

Fuentes, A. (PhD-student)*, Lüdtke, O. (Supervisor), Robitzsch, A.

Leibniz Institute for Science and Mathematics Education, Germany

*Presenting author

### Suggested talk duration

15-20 minutes

### Summary

Using observational data, the causal effect of a treatment on some outcome can be estimated without bias if the estimation controls for all confounders, that is, covariates that are associated with both treatment assignment and the outcome. This can be particularly challenging in multilevel settings, where covariates are typically numerous and located at different levels of analysis. Hong and Raudenbush (2003) first extended the potential-outcomes framework of Rubin (1978) to multilevel causal analysis, and in the last 15 years a literature has emerged on the properties of propensity score (PS) methods that can adjust for a large set of covariates when the data are clustered and treatment is administered at level 1.

In the present paper, we conduct Monte Carlo simulations to assess the performance of various PS adjustment strategies, in terms of the bias and variance of the estimates they produce. We compare well-known PS matching, weighing and stratification estimators in various sample size conditions (number of clusters, and units per cluster), and with different intraclass correlations for the treatment and outcome variables.

More specifically, the simulations were designed to evaluate and expand two key findings of the multilevel PS literature.

First, we show that a correctly specified PS can remove a sufficient amount of confounding bias on its own, that is, when an outcome model is absent or misspecified. The importance of this lies in the fact that outcome models have consistently been shown to have a greater influence than treatment models in doubly-robust estimation procedures (e.g., Su, 2008; Li, Zaslavsky, &

Landrum, 2013). Our results indicate that outcome models are indeed more influential, but also that the benefits of modeling treatment assignment are not trivial.

Second, we show how shrinkage distorts the information of PSs estimated with random-effect models, and causes them to underperform. Previous studies have concluded that PSs estimated with group indicators (i.e., fixed-effects) are superior in small samples (e.g., Thoemmes & West, 2011). Our simulations confirm this finding and provide further guidance for modeling the PS.

Throughout, we highlight the modelling choices that practitioners face, and the findings that should guide those choices.

## Relevance to conference theme

The topics of this talk are multilevel extensions of well-established methodologies. Attendees with some knowledge of propensity score methods will benefit the most, but anyone with a multilevel background should be able to follow.


## Keywords

Propensity Scores, Causal Inference.


## References


Hong, G., & Raudenbush, S. W. (2003). Causal inference for multi-level observational data with application to kindergarten retention study. *2003 Joint Statistical Meetings*. American Statistical Association.

Li, F., Zaslavsky, A. M., & Landrum, M. B. (2013). Propensity score weighting with multilevel data. *Statistics in Medicine*, 32(19), 3373-87.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, 6(1), 34-58.

Su, Y. (2008). Causal inference of repeated observations: a synthesis of the propensity score methods and multilevel modeling. Paper presented at the PolMeth XXV, 25th Annual Summer Meeting, University of Michigan, Ann Arbor.

Thoemmes, F. J., & West, S. G. (2011). The use of propensity score for nonrandomized designs with clustered data. *Multivariate Behavioral Research*, 46(3), 514-543.

# Modeling longitudinal dyadic data in the SEM framework

Gistelinck, F. (PhD-student)[1*], Loeys, T. (supervisor)[1]

[1] Ghent University, Belgium

* Presenting author

**Suggested talk duration (15-60 minutes)**

## Summary (max. 500 words)

In social and behavior science, researchers may be interested in the effect of an antecedent on the current behavior or emotional status of a person. However, as daily lives are rarely spent in isolation, close dyadic relationships are nowadays more often examined instead of individuals. Consequently, the statistical framework for dyadic research flourished over the last decade. One of the most widely used models within that area is the actor-partner interdependence model (APIM), which models the effect of a predictor measured across dyad members on one's own and one's partner outcome. When such dyadic data are measured repeatedly over time, the necessity of a good statistical model becomes even more profound.

Indeed, there are a lot of statistical challenges when addressing such research questions concerning longitudinal dyadic data. First, one needs to take the non-independence between the two members of a dyad into account. As the two members of a dyad can be considered more similar or dissimilar than two random people, one might wonder to what extent the time-average emotional status of the two members of a dyad are correlated with each other. Second, as one is dealing with a repeated measurement-design, one must also consider the non-independence of the observations within a dyad member. How strong is the association of an outcome on one day with the outcome on the next day within a given dyad member? Third, one should allow for the effect of one person's behavior or emotions on his/her own score (i.e., actor effect) as well as for the effect of one partner's behavior or emotions on that person's score (i.e., partner effect). Fourth, one should differentiate the explained variation on within-dyad (member) level from the variation on between-dyad (member) level. Consequently, a distinction should be made on time-averaged and time-specific effects. In

order to address these four issues, an extension of the APIM will be introduced that allow researchers to analyze longitudinal dyadic data.

When considering the implementation of this complex model, further challenges arise. For instance, the statistical software should be able to handle between-dyad variation at level-2 (the so-called G-side), while simultaneously allowing for level-1 residual covariance structures (the so-called R-side). The latter can take very complex forms as it inhibits both the correlation within a dyad at a specific time point and the correlation for the measurements over time within each dyad member (i.e., the autocorrelation), which is often depicted as a first-order autoregressive process. Up until now, only a few multilevel modeling software packages (such as *proc mixed* in SAS) are able to readily fit this longitudinal extension of the APIM, also called the L-APIM. We will show how the model can be implemented in structural equation modeling (SEM) software, such as the R-package *lavaan*. Moreover, a Shiny-application was developed to enable applied dyadic researchers to fit the L-APIM on their longitudinal dyadic data within the SEM framework.

## Relevance to conference theme

The talk focusses on a specific multilevel model in dyadic research and its issues when applied in standard multilevel modeling software. These issues will be explained and an alternative implementation in the structural equation modeling framework will be introduced. An user-friendly shiny-app was developed allowing applied researchers to fit this advanced model.

## Keywords (max. 3)

Longitudinal dyadic data, structural equation modeling, multilevel modeling

# Multiple imputation of missing data in multilevel models with random slopes and nonlinear effects

Grund, S.[1,2*], Lüdtke, O.[1,2], Robitzsch, A.[1,2]

[1] Leibniz Institute for Science and Mathematics Education, Germany

[2] Centre for International Student Assessment, Germany

* Presenting author

**Suggested talk duration (15-60 minutes)**

15 - 20 minutes (research talk)

**Summary (max. 500 words)**

Missing data are a pervasive problem in multilevel research. Moreover, the treatment of multilevel missing data with methods such as *multiple imputation* (MI) is often challenging because they require that the multilevel structure of the data and the features of the substantive analysis model are taken into account. This is particularly true if the substantive analysis model includes random slopes or interaction effects (e.g., cross-level interactions) because these effects imply complex, nonlinear associations between variables that are difficult to address with conventional methods for multilevel MI.

In recent years, it has been argued that *substantive-model compatible* MI (SMC-MI) can provide an effective treatment of missing data even in complex multilevel analyses by ensuring that imputations remain consistent with the substantive analysis model. Several different methods adopting some form of SMC-MI can now be found in the literature and are currently starting to become widely available in statistical software.

In the present talk, we provide a comparison between these methods on the basis of both theoretical considerations and the results of a series of simulation studies. In this context, we consider applications of multilevel models with random slopes, interaction effects, models with centered

covariates and interactions at both level 1 and 2, and applications with more than one substantive analysis model (e.g., multiple outcomes of interest). Based on our findings, we discuss the strengths and weaknesses of the methods considered and provide recommendations for research practice.

### Relevance to conference theme

The talk is relevant to both methodological and applied researchers working with multilevel models. From a methodological perspective, the talk offers an in-depth discussion of the procedures that have been proposed to deal with nonlinear effects (e.g., random slopes, interactions) in multilevel models with missing data. From an applied perspective, the talk offers an evaluation of the current procedures and recommendations for practice.


### Keywords (max. 3)

multilevel models, missing data, multiple imputation

# Multilevel latent class (MLC) modelling of simulated upper-level causal effects in observational data

Harrison, WJ.[1,2*], Baxter, PD.[2] Gilthorpe, MS.[1,2,3]

[1] Leeds Institute for Data Analytics, University of Leeds, Leeds, LS2 9NL, UK

[2] School of Medicine, University of Leeds, Leeds, LS2 9LU, UK

[3] The Alan Turing Institute, London, UK

* Presenting author

**Suggested talk duration (15-60 minutes)**

20 minutes

**Summary (max. 500 words)**

Background

Observational data may be structurally complex, with connections between information collected at different levels of a hierarchy; and covariates at any level may affect outcomes. We consider an example, based on routinely collected cancer-registry data, where patients are clustered within healthcare providers. Patient population characteristics may vary across providers (termed 'casemix'), which may lead to differences in patient outcomes. To assess potential causal effects at the provider level, patient casemix must first be balanced across providers.

We propose the use of multilevel latent class (MLC) modelling to partition modelling strategies across the hierarchy. We model prediction at the lower level (to accommodate patient casemix differences) and model causal inferences at the upper level (to assess provider-level causal effects).

Data simulation

Data are simulated, with a homogeneous lower-level group, independent binary and continuous upper-level covariate effects, and a continuous outcome. Both the data structure and the distribution of lower-level covariates are based on values sourced from a real-world dataset. Unique sets of 100 simulated datasets are generated, using a range of coefficient effect values, error variances and simulation seeds.

Modelling

The modelling strategy determines upper-level latent classes based on differences in lower-level characteristics, ensuring these upper-level classes are balanced with respect to lower-level casemix. Residual outcome differences are then due to covariate effects operating at the upper level, which are simulated within these data. Interest lies in the recovery of the upper-level covariate effects, as designed into the data simulations.

The binary and continuous covariate effects are analysed separately. Multiple simulated datasets are similarly modelled, generating median recovered values and credible intervals for each simulated upper-level coefficient. Two upper-level latent classes are required as a minimum, to distinguish outcome differences.

Results

Models contained one lower-level latent class and up to five upper-level latent classes. For the binary upper-level covariate, results were consistent across models, and median recovered values were almost identical to simulated coefficient values throughout. For the continuous upper-level covariate, median recovered values improved as the number of upper-level classes were increased; all estimates were within credible intervals for models with three latent classes or more.

For both upper-level covariates, credible intervals widen with increased error variance. For the continuous upper-level covariate, they also widen as the simulated upper-level coefficient value is increased.

Discussion

The MLC modelling approach achieves successful recovery of parameter values for both binary and continuous upper-level covariates (for a continuous outcome). Very small simulated values of the upper-level coefficients were not recovered as well as higher values, which may be due to the variability introduced during simulation dominating the coefficient parameter value. There is also some attenuation of effect seen for the continuous upper-level covariate.

Modelling for prediction and for causal inference are partitioned across the hierarchy, an approach only feasible through use of latent variable methodologies. There is much scope to extend the assessment of upper-level causal effects by consideration of a multivariable DAG.

**Relevance to conference theme**

Innovative methodological approach using multilevel modelling in a latent class framework

**Keywords (max. 3)**

Latent class

Causal inference

Casemix

# Using R to Evaluate Collinearity in Mixed Models.

John Hendrickx

Senior statistical programmer at Nutricia Research, Utrecht, the Netherlands

**Suggested talk duration (15-60 minutes)**
30 minutes

**Summary (max. 500 words)**
Strong correlations among the independent variables in a model can lead to unstable results due to large standard errors for the estimates. There are well documented methods for evaluating the impact of collinearity in linear regression models but these methods are geared to models with continuous variables only. As an alternative, the R package “perturb” can be used to evaluate the impact of collinearity by adding random noise to selected variables. This method can deal with transformations, interaction effects, categorical variables. The perturb package has been enhanced to work with the nlme and lme4 packages for mixed models in R. This makes it suitable for collinearity with a random slope variable or sparse categories in a levels variable.

**Relevance to conference theme**
Collinearity is not a topic that is commonly associated with mixed models. For some models, small changes in the data could mean substantial changes in the results or estimation issues. The paper contains an interesting example of this.

An earlier version of this paper was presented at the PhUSE EU Connect conference in Frankfurt, Nov. 4-7 2018. This version will also examine perturbations and posthoc tests (“contrasts” in SAS terminology)
https://www.lexjansen.com/phuse/2018/as/AS03.pdf

**Keywords (max. 3)**
Collinearity, simulation, sparse levels

**Modeling Student Mobility Using Hierarchical Networks**

Johnson, T. L.[1*], Sweet, T. M.[1]

[1] University of Maryland, College Park, USA

* Presenting author, PhD-student. Doctoral advisor: Gregory R. Hancock. Project supervisor: Tracy M. Sweet

**Suggested talk duration (15-60 minutes)**

15-20 minutes

**Summary (max. 500 words)**

## Introduction

Student mobility, or school transfers unrelated to academic promotion (Rumberger, 2003), has long been recognized as a risk factor for adverse educational outcomes, such as dropout and low educational performance among students and low morale and higher administrative burdens among teachers and staff (Gruman, Harachi, Abbott, Catalano, & Fleming, 2008; Rumberger, 2003; Rumberger & Larson, 1998). Examination of peer social networks suggests that contextual effects of mobility exist such that even non-mobile students from schools with high rates of mobility are at risk (South, Haynie, & Bose, 2007). Furthermore, patterns of school mobility may be heterogeneous. In an urban setting, similar schools—that is, schools serving populations with similar demographic profiles—tend to form clusters based on student transfer (Kerbow, 1996); however, this may not be the case in rural or suburban settings.

Though the impact of mobility appears to be profound at both the student and school level, less is known about the inter-school network process of student transitions outside of urban contexts. To bridge this gap, the current study implements a novel application of hierarchical social network methods to model student mobility across networks of schools. We examine factors related to the formation of these multilevel networks among high schools within counties using data from a state longitudinal data system.

## Methods & Results

Previous work has modeled student mobility as an outcome or predictor based on presence of a move or number of moves (Engec, 2006; South, Haynie, & Bose, 2007). Alternately, multiple membership models, which do not directly model mobility but rather allow student outcomes to be affected by multiple school-level units, have gained attention in recent years (Smith & Beretvas, 2017). The current study utilizes hierarchical latent space network models (Sweet, Thomas, & Junker, 2013), which model the likelihood of observing a specific network out of the set of all possible networks, to examine student mobility across schools. This modeling framework allows us to account for factors that may exist in the school-student transfer process like the clustering structure identified among urban schools (Kerbow, 1996).

The present model examines within-county networks where nodes are schools and ties are students moving between schools. Between-county transfers are not modeled. School-level variables include unauthorized absences, pass rates on state exams, and average wages of student workers. County-level variables include information on annual educational expenditures and median income. Across programs, the rate of students enrolled in free and reduced price lunch (FRL) programs had a large, significant effect on the probabilities that a school would send or receive a student within a given year.

Our use of hierarchical network analysis to model student mobility represents an advancement in understanding county- and school-level factors that affect the formation and reinforcement of mobility networks. As state longitudinal data become more widely available, studies of this nature may help us understand mobility networks across settings, better informing interventions and policies related to student retention.

**Selected References**

Engec, N. (2006). Relationship between mobility and student performance and behavior. *The Journal of Educational Research, 99,* 167-178.

Gruman, D. H., Harachi, T. W., Abbott, R. D., Catalano, R. F., & Fleming, C. B. (2008). Longitudinal effects of student mobility on three dimensions of elementary school engagement. *Child Development, 79,* 1833-1852.

Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk, 1,* 147-169.

Rumberger, R. W. (2003). The causes and consequences of student mobility. *Journal of Negro Education, 72,* 6-21.

Rumberger, R. W. & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education, 107,* 1-35.

South, S. J., Haynie, D. L., & Bose, S. (2007). Student mobility and school dropout. *Social Science Research, 36,* 68-94.

Smith, L. J. W. & Beretvas, S. N. (2017). A comparison of techniques for handling and assessing the influence of mobility on student achievement. *The Journal of Experimental Education, 85,* 3-23.

Sweet, T. M., Thomas, A. C., & Junker, B. W. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics, 38*(3), 295-318.

**Relevance to conference theme**

Hierarchical network models utilize the familiar framework of multilevel modeling to estimate the effect of a set of covariates on the observation of a given network when networks are not randomly sampled within the population. Our advanced application of multilevel network models to student mobility is novel within an education setting and can inform future work within the contexts of analyzing mobility and of advancing multilevel methodology.

**Keywords (max. 3)**

Latent space network analysis, student mobility, hierarchical models

# (Not) Everybody Does: Testing for Individual Differences and Similarities in Hierarchical data

Klaassen, F.[1*], Rouder, J.N.[2,]

[1] Utrecht University, the Netherlands

[2] University of California, Irvine, USA

* Presenting author – PhD-student. Supervisor: prof. H. Hoijtink (h.hoijtink@uu.nl)

**Suggested talk duration**

20 minutes

**Summary (370 words)**

Psychological experiments often require participants to repeat many trials in several experimental conditions. The resulting nested datasets are often analyzed by means of mixed models. Theories can predict order constrained differences between experimental conditions.

Consider the Navon letter task, where participants are presented with a large letter composed of small letters, that are either congruent (same letter) or incongruent (different letter), and asked to report the shape of the large, or the shape of the small letter. It can be expected that the Response Time (RT) for congruent trials is shorter than for incongruent trials. These order constraints can be evaluated by means of Bayesian hypothesis testing. The constraints can be imposed on the average-effect level, that is: the average RT in incongruent trials is longer than the average RT in congruent trials.

Some theories predict order constraints between conditions to hold not only on average, but also across individuals. The constraints could be imposed at the individual-effect level, that is: the RT in incongruent trials is *for all individuals* longer than the average RT in congruent trials.

The goal of this research is to evaluate the possibility that the average effect is (near) zero, but individuals differ in the direction of their effects. An

example would be asking participants to throw a ball with left and right hands, and measure the distance. On average, the right-handed throws will be further than the left-handed throws, because there are more right-handed people in the population. However, at an individual level, we find people with further left-throws (lefties) and people with stronger right-throws (righties). In this research we evaluate whether the effect of the Navon letter task is omnipresent, that is, does everybody indeed have the same local or global preference?

This talk presents three adaptations of the Navon letter task designed to evoke different outcomes. The task was manipulated such as to create a condition where the global orientation would be preferred for everybody, a condition where the local orientation would be preferred for everybody, and a condition where the average effect would be around zero, and individuals differ in whether they present global or local preference. To analyze this question, the methodology is evaluated by means of a brief simulation.

**Relevance to conference theme**

New perspective to hypothesis testing for individual effects in hierarchical data. Two different approaches are presented and compared to group-effect hypothesis testing: case-by-case analysis with evidence synthesis and placing order constraint on the hierarchical structure.

**Keywords (max. 3)**

Bayesian hypothesis testing; order constraints; evidence synthesis

# Calculating intraclass correlation coefficients in multilevel models for count responses

Leckie, G.[1*]

[1] School of Education, University of Bristol, UK

* Presenting author

**Suggested talk duration (15-60 minutes)**

**Summary (max. 500 words)**

A standard step when fitting multilevel models to continuous responses is to calculate the degree of clustering in the response using the intraclass correlation coefficient (ICC). When fitting multilevel models to binary and ordinal responses, an analogous ICC can be calculated via the latent response formulation of these models. However, when fitting multilevel models to count responses, there is no easy way to calculate ICC statistics and many applied researchers fail to report the degree of clustering in their analyses. A simulation approach has been proposed but is computationally intensive and somewhat complex to implement. In a recent publication, we drew attention to a little-known existing ICC formula for the special case of a two-level random-intercept Poisson model. In this talk, we show how this approach naturally extends to models with additional levels, random coefficients, and more flexible negative binomial models which allow for overdispersion in the counts, a phenomenon which often occurs in practice. We confirm the formulae give the same results as the simulation approach and we illustrate their utility with an application to studying student absenteeism in secondary schools in England.

**Relevance to conference theme**

**Keywords (max. 3)**

Intraclass correlation coefficients; count response models

# The social relations model for count data: To Bayes or not to Bayes

Loncke, J.[1*], Loeys, T.[2]

[1] Ghent University, Belgium (PhD-student)

[2] Ghent University, Belgium (Supervisor)

* Presenting author (also indicate if the presenting author is a PhD-student by adding the text 'PhD-student' and add the name of your supervisor)

## Suggested talk duration (15-60 minutes)

15 minutes

## Summary (max. 500 words)

The family social relations model (SRM) is widely used to identify the sources of variance in interpersonal dispositions in families. Traditionally, it makes use of dyadic measurements that are obtained according to a round-robin design, where each family member rates the other family members on a specific interpersonal disposition. In this study, we will consider two challenges with the family SRM. A first challenge concerns the data design. Sometimes family researchers are interested in family dynamics that are based on only a particular subset of relationships, e.g. parent-adolescent interactions. To increase the efficiency of data collection, the dyadic measurements can then be obtained from a block design. Therefore, we will adapt the family SRM to data that are obtained from a block design, which is restricted to merely intergenerational dyadic measurements. A second challenge lies in the nature of the dyadic measurements. Typically the dyadic measurements are assumed to be normally distributed, however family researchers may also be interested in non-normal family data. For example, they may be interested in the family dynamics behind the number of different activities that a family member reports with another family member. If they then want to identify the sources that explain the most variability in perceived co-activity between families, the family SRM needs to be accommodate to count data. To model hierarchical count data, one can make use of two frameworks: structural equation modeling (SEM) framework and multilevel framework. These two frameworks result in equivalent models of which the parameters are usually estimated

using a maximum likelihood (ML) estimator. However, it is known that there are limitations to the ML estimator. For instance, the ML variance estimators might be biased in small samples. A Bayesian approach using Gibbs sampling could overcome the shortcomings of the ML estimator. However, it has also been shown that modeling the SRM without family roles for normal dyadic measurements in the Bayesian approach might result in biased estimators for the variances for small cluster sizes in combination with small sample sizes. In this presentation we evaluate whether the Bayesian approach for the family SRM, which involves a small group size, becomes even more problematic for count data? And if so, is the ML estimator then a better alternative? These questions are answered by means of a simulation, in which the performance of the Bayesian estimator in the multilevel framework is compared to the performance of the ML estimator in the SEM-framework. As an illustration we consider intergenerational co-activity data and contrast family dynamics between non-divorced families and stepfamilies using the SRM.

**Relevance to conference theme**

The presentation discusses the family SRM, a multilevel model that is used to study relational or dyadic data from multiple persons in families. Specifically, it explores which estimation technique provides the best performance when applying the model to count dyadic measurements.

**Keywords (max. 3)**

Structural equation modeling, family social relations model, Bayesian modeling

# A joint modelling approach to relate within-individual variability in a repeatedly measured exposure to a future outcome, allowing for measurement error in the repeated measures

Parker,R.M.A.[1*], Goldstein,H.[2], Heron,J.[1], Howe,L.[1], Leckie,G.[2], Tilling,K.[1]

[1] Population Health Sciences, Bristol Medical School, University of Bristol, UK

[2] Centre for Multilevel Modelling, School of Education, University of Bristol, UK

* Presenting author

### Suggested talk duration (15-60 minutes)

20-30 minutes

### Summary (max. 500 words)

Analyses seeking to relate repeatedly measured exposures to a future health outcome have often treated within-individual variability in the exposure as a nuisance that should be removed or ignored, with such variation being viewed as a consequence of normal biological variability and/or measurement error. Within-individual variability in the exposure might be an important risk factor to consider in its own right, however: for example, within-individual variability in blood pressure (BP) is an independent cardiovascular risk factor above and beyond mean BP.

Using data from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort study, we propose a Bayesian joint modelling approach to relate within-individual variability in blood pressure to a later indicator of cardiovascular health (left ventricular mass). The structure of the resulting (multilevel) submodel for the repeatedly-measured exposure consists of replicates within occasions within individuals, allowing short-term within-occasion variation ("measurement error") to be distilled from the longer-term, between-occasion, variation which may be of greater biological interest. The between-occasion variation is modelled as a function of relevant covariates as well as of a random effect which allows each individual to have more, or less, variation than that implied by the other covariates in the model. This random effect is then related to the distal outcome within the same model, by including it in a covariance matrix alongside the residual variance from the distal outcome, or as a predictor for that outcome.

Unlike two-stage methods to relate within-individual variability to a distal outcome – in which a summary measure of variation for each individual is calculated and then included as an exposure in a separate model of the distal outcome – this joint modelling approach takes account of the precision with which the within-individual variation was estimated. In addition, the structure of the model allows shorter-term variation ("measurement error") to be distinguished from the longer-term variation, with the latter related to a later health outcome in a manner which better reflects biological interest.

## Relevance to conference theme

Multilevel modelling

## Keywords (max. 3)

longitudinal data analysis; within-individual variation; joint-modelling.

# Multiple imputation and selection of ordinal level-2 predictors in multilevel models: analysis of the relationship between student ratings and teacher beliefs and practices

Grilli, L.[1], Marino, M.F.[1], Paccagnella, O.[2], Rampichini, C.[1*]

[1] Department of Statistics, Computer Science, Applications 'G. Parenti', University of Florence, Italy
[2] Department of Statistical Sciences, University of Padua, Italy
* Presenting author

**Suggested talk duration**
20 minutes

**Summary (max. 500 words)**
We devise a strategy to handle ordinal level-2 predictors of a two-level random effect model in a setting characterized by two nontrivial issues: (i) level-2 predictors are severely affected by missingness; (ii) there is redundancy in both the number of predictors and the number of categories of their measurement scale. We tackle the first issue by considering a multiple imputation strategy based on information at both level-1 and level-2. For the second issue, we consider regularization techniques for ordinal predictors, also accounting for the multilevel data structure. The work is motivated by a case study at the University of Padua about the relationship between student ratings of a course and several characteristics of the course, including teacher feelings (ordinal predictors) and practices (binary predictors) collected by a specific survey with nearly half missing respondents.

**Relevance to conference theme**
Advanced applications aspect of multilevel modelling.

**Keywords**
Lasso, MICE, University course evaluation

# Multiple Imputation in Three-level Models

Menon, N.[1], Richardson, A.M.[1*]

[1] National Centre for Epidemiology & Population Health, Australian National University, Australia
* Presenting author

**Suggested talk duration (45 minutes)**

**Summary (max. 500 words, 358 at the moment)**
Biostatistical methodologists are directing much attention at the issues surrounding multiple imputation of missing values in multilevel data. The field of methods for single level and two level data has matured considerably. Modellers can choose from the fully conditional specification (FCS) of van Buuren et al, and the joint modeling approach of Carpenter & Kenward, amongst others. Recent years have seen the introduction of packages in R (mice and jomo) for each of these, with several extensions to mice also now available.
However three level data remains an open problem. Gelman & Hill have suggested a two-step process for imputing in two-level models which could be extended to three level models. Andridge has suggested treating the top level as fixed, then proceeding as for two-level models.
Another open problem is the most efficient way to estimate derived variables e.g. quadratic terms, interactions, or contextual variables such as a group mean or deviations from a group mean. Contextual variables are particularly important in a variety of applications and a variety of methods have been proposed for imputing these. Treating the contextual variable as "just another variable" is a simple option. There are also approaches which more closely mirror the relationship of the contextual variable to its components, such as the rejection sampling approach of Bartlett & White, and the passive imputation implemented in the FCS approach mentioned previously.
In this talk we will compare the two-step process and the fixed effect approach to complete case analysis for a three-level model with missing data at level 1, level 2, and both level 1 and level 2. Missingness either completely at random or at random will be generated at the rates of 0, 20% and 50%. Performance measures such as convergence rates, mean square error, coverage and bias will be compared across the $3 \times 3 \times 2 \times 3 = 54$ scenarios.
The talk will also be illustrated with data from population health research in Australia. Selection will be made from a cluster randomized trial of smoking cessation, a retrospective analysis of laboratory prediction of hepatitis C, and an ecological study of risk factors associated with HIV infection in India.

**Relevance to conference theme**
This talk fits the conference theme of innovative applications and software. The capabilities of R software will be explored in a complex three-level structure. The methods will be applied to data from Australia.

**Keywords (max. 3)**
Three-level data; simulation study; population health.

# Why country dummies sometimes do not do the job. How to get the within-estimator of cross-level interactions with pooled cross-sections.

Schmidt-Catran, Alexander W.[1*], Giesselmann, Marco[2,3]

[1] Institute of Sociology – Goethe University Frankfurt, Germany
[2] DIW Berlin, Germany
[3] Institute of Sociology – Bielefeld University, Germany
* Presenting author

**Suggested talk duration: 20 minutes**

**Summary (max. 500 words)**
Multilevel models with persons nested in countries are increasingly popular in cross-country research. Recently, social scientists have started to analyze data with a three-level structure: persons at level 1, nested in year-specific country samples at level 2, nested in countries at level 3. By using a country fixed-effects estimator, or an alternative equivalent specification in a random-effects framework, this structure is increasingly used to estimate within-country effects in order to control for unobserved heterogeneity at the country level. For the main effects of country-level variables, such estimators have been shown to have desirable statistical properties (Fairbrother 2014). However, estimators of cross-level interactions in these models are not exhibiting these attractive properties: as algebraic transformations show, they are not completely independent of between-country variation—they carry country-specific effect heterogeneity. Monte Carlo experiments consistently reveal the standard approaches to within estimation to provide biased estimates of cross-level interactions in the presence of unobserved correlated moderators at the country level. To obtain an unbiased within-country estimator of a cross-level interaction, effect heterogeneity must be systematically controlled for. We propose three alternative model specifications that do this job. We demonstrate the relevance of our claim by replicating a published analysis.

*Literature:*

Fairbrother, Malcolm. 2014. “Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets.” Political Science Research and Methods 2(1):119–40.

**Relevance to conference theme**
We propose an extension of a common three-level multilevel specification, which is typically applied to pooled cross-sectional international survey data.

**Keywords (max. 3)**
Within-estimation, pooled cross-sections, effect heterogeneity.

# Income Equality in Achievement among US Elementary Schools: A Random Coefficients Model with Data MAR

Shin, Y.[1*], Raudenbush, S.W.[2]

[1] Virginia Commonwealth University, USA
[2] University of Chicago, USA

**Suggested talk duration (15-60 minutes)**
40 minutes

**Summary (max. 500 words)**
Increasing inequality in educational outcomes has recently captured the attention of social scientists. In particular, parental income has become an increasingly important predictor of educational achievement. Educational sociologists have long noted that the effectiveness of policies for reducing inequality may depend on how inequality is distributed between and within schools: a) Schools serving high-income children may have much higher achievement levels than do schools serving low-income children; b) Within-school differences in family income drive within-school differences in achievement. However, within-school disparities may differ substantially from school to school. Even in a society characterized by substantial inequality, some schools may be remarkably egalitarian. A popular model for describing the multiple sources of inequality is a hierarchical linear model that incorporates school-mean differences in family income and school-specific regression coefficients. However, two methodological challenges confront this kind of analysis. First, missing data, particularly on income, is substantial in large-scale surveys. Second, the average income of parents in a school must be estimated from the sample, generating an errors-in-variables problem. Third, the average income may moderate the effect of parental income. Using nationally representative survey data on US elementary schools, we address these challenges within the framework of maximum likelihood by an imputation model for the joint distribution of the outcome and covariates that may have random coefficients. With data assumed missing at random, we estimate the non-standard imputation model by the EM algorithm using Adaptive Gauss-Hermite Quadrature (AGHQ). To avoid the "curse of dimensionality" that can afflict AGHQ, we apply iterated expectations to a uniquely factored complete-data log-likelihood. We demonstrate the accuracy of the approach by simulation, and compare our results to those obtained using standard approaches to the mixed model. The novel approach enables us to provide the most complete description to date of income inequality in achievement among US elementary schools.

**Relevance to conference theme**
methodological aspect of multilevel modelling, and advanced application

**Keywords (max. 3)**
errors-in-variables; multiple imputation; auxiliary variables.

# Optimal developmental trajectory group analyses: Which parameters should (not) be constrained to accurately estimate growth mixture models?

Sijbrandij, J.J.*[1] (PhD student, primary supervisor Bültmann, U.),
Hoekstra, T.[1], Almansa J.[1], Peeters, M.[2], Bültmann, U.[1], Reijneveld, S.A.[1]

1 Department of Health Sciences, Community & Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
2 Department of Interdisciplinary Social Science, Utrecht University, Utrecht, The Netherlands

**Suggested talk duration (15-60 minutes)**
30 minutes

**Summary (max. 500 for Utrecht, now 462)**
Growth mixture modeling (GMM) has become a standard approach in grouping people based on their development over time. Convergence issues and impossible values (e.g. negative variances) appear frequently in GMM, especially for smaller sample sizes. To solve these issues, researchers often limit the number of estimated parameters by constraining either random effect variances to be equal over classes or residual variances to be equal over time or equal between classes. Constraining these variances to be equal, when they in fact differ, can lead to biased estimates. Yet, it remains unknown which variances can best be constrained to obtain the least biased estimates.

The aim of this study is to determine which variances are best to constrain in GMM to (1) obtain the least biased estimates, (2) the most accurate assignment of individuals to developmental trajectory classes, (3) and highest rate of estimated classes corresponding with simulated classes.

A simulation study was conducted, followed by an illustration with empirical data of the Tracking Adolescent Individuals' Lives Survey (N = 2,227), a population-based cohort of Dutch pre-adolescents followed into young adulthood over six measurement waves. In the simulation study,

we compared models that differed in variance constraints, sample size (100, 300 or 1000) and the distance between the classes.

The model constraining random effect variances and residual variances over classes performed worst. For a sample size of 300 or larger, the unrestricted model and the model which constrained random effect variances over classes and residual variances over time but unconstrained across classes performed best. For a sample size of 100, this model with constrained random effect variances across classes and constrained residual variance over time performed best. For a larger distance between the classes, the same models performed best, although the difference between the best models (unconstrained model and model which constrained random effect variances across classes and residual variance over time) and the worst model (random effect variances and residual variances constrained across classes) model was smaller.

To conclude, sample size is an important factor in finding the most appropriate model. In general, models that constrained residual variances to be equal over time performed better than models that constrained other variance parameters. Therefore, if some parameters need to be constrained to aid model convergence or solve impossible values, we recommend starting with constraining the residual variance over time. We discourage to start with constraining the residual variance to be equal across classes. The model specification we recommend differs from the default specification in commonly used software packages. For example, Mplus constrains random effects and residual variances to be equal across classes, if the data is read in in wide format. It is therefore very important that researchers carefully consider possible constraints of residual variances and random effect variances, rather than following the software's default specifications.

**Relevance to conference theme**

The abstract addresses the specification of level 1 and level 2 variances and their effects on the parameter estimates in growth mixture modeling. This is an important issue in multilevel modeling, which can strongly affect model results.

**Keywords (max. 3)**

Growth mixture modeling, convergence, residual variance

# Assessing Individual Change Processes

## Bayesian Covariance Structure Modelling for negative associations among patients with personalized treatments


Wouter Smink[1,2*†], MSc

prof. dr. ir. Jean-Paul Fox[2]

dr. Anneke M. Sools[1]

dr. Erik Tjong Kim Sang[3]

dr. Ben L. de Vries[3]

prof. dr. ir. Bernard P. Veldkamp[2]

prof. dr. Gerben J. Westerhof[1]

[1] Department of Psychology, Health & Technology

[2] Department of Research Methodology, Measurement & Data Analysis;

University of Twente, Enschede, The Netherlands

[3] Netherlands eScience Centre (NWO); Amsterdam, The Netherlands

* Presenting and corresponding author, w.a.c.smink@utwente.nl

† PhD-student supervised by prof. dr. Gerben J. Westerhof

prof. dr. ir. Bernard P. Veldkamp

dr. Anneke M. Sools

prof. dr. ir. Jean-Paul Fox



Submitted to the *International Multilevel Conference* multilevel@uu.nl, 15-01-2019

**Keywords**

- Personalized treatments
- Bayesian Covariance Structure Modelling (BCSM)
- Negative Clustering Effect

**Suggested duration**

- talk 20 minutes
- discussion 10 minutes

**Potential questions for discussants**

- Are there alternatives methods for studying *personalized treatments*?
- How can there be *negative variance*?
- How are *negative variance* and *personalized treatments* related?

**Preferably in the same track**

- Fox & Smink (2019). Bayesian Covariance Structure Modelling: A simulation study combined with real data to present novel method for multilevel model.

## Relevance to conference theme

Our work aims to advance *Therapeutic Change Process Research* (TCPR), a field that aims to relate the *individual* in-therapeutic change processes to the (post-therapeutic) outcomes of interventions. As the study of individual change usually relies on repeated measurements of individuals (cf. Raudenbush & Bryk, 2002), progress in the field of TCPR is intricately tied to (methodological) breakthroughs in multilevel modelling.

As people differ widely in how they respond to therapy, *individualized interventions* mandate methods able to assess how individuals changed through therapy. Yet, most statistical models express the treatment effect as an average over a group of individuals. By showing how we model *personalized change*, we expect that we will interest psychotherapy researchers, methodologists and researchers coming from various disciplines: are not all researchers by some degree interested in the question *what* treatment, by *whom*, is most effective for *this* individual with *that* specific problem (Paul, 1967)? As the *International Multilevel Conference* (IMC) covers statistical and methodological aspects of multilevel modelling, we feel that the IMC hosts an exceptionally broad audience to showcase our novel method.

We rely on *Bayesian Covariance Structure Modelling* (BCSM) to study individual change. Preferably, this talk is therefore programmed after the talk of Fox and Smink (2019). We introduce the BCSM in comparison to standard multilevel models by elaborating on the estimation and interpretation of change processes and (clustering) effects of counsellors. We rely on intuitive and realistic examples coming from a (simple) simulation study and real-data from Lamers, Bohlmeijer, Korte, and Westerhof (2015). Both examples will be aimed at enhancing the understanding of the advantages modelling change processes through BCSM.

**Summary**

*Therapeutic Change Process Research*
The field of Therapeutic Change Process Research (TCPR) aims to identify the mechanisms through which treatments attain psychotherapeutic change. At its core, many TCPR research questions pertain to the question: *what* treatment, by *whom*, is most effective for *this* individual with *that* specific problem (Paul, 1967)? As almost all studies are aimed at demonstrating average group-level effects, there is a discrepancy between research and practice: counsellors treat individuals, but they only know how to treat groups.

*Multilevel models for studying change*
Many individual change phenomena can be represented through a two-level hierarchical model (Raudenbush & Bryk, 2002), which is why multilevel models arise quite naturally for studying individual change. The first level represents each client's development by an individual growth trajectory that depends on the repeated measures for each client; the second level unit represents variables that are not repeatedly measured, such as gender, income, or depressive symptoms. As counsellors (almost) always treat more clients, clients could be viewed as grouped within their counsellor.

*Negative clustering effects*
As people differ widely in how they react to events, properly modelling the heterogeneity might provide a key avenue for modelling individual effects. In such a study of change, the influence of each counsellor is often ignored. In case of a personalized treatment however, **increased heterogeneity** can be expected among individuals over time: the counsellor will succeed better in personalization for some than for others. The effectiveness of the personalized treatment given by the therapist differs across individuals, where some individuals benefit more from the therapist than others.

This can lead to **negative associations** among measurements across individuals. Traditionally, the positive association between measurements is modelled through a random effect with a *positive variance*. To model the negative associations between measurements, one would need to entertain the possibility of modelling random effects with **negative variances**.

*Negative variance of random effects*
In a multilevel modeling approach, a random effect is used to model dependencies among treated individuals. However, the random effect

variance is restricted to be positive and, as a result, implies a positive association among individuals. Negative associations among measurements caused by a therapist, who increases the heterogeneity among treated individuals, would require the modelling of a **negative random effect** variance.

*Bayesian Covariance Structure Modelling* (BCSM)
This apparent contradiction requires a different multilevel modeling parameterization, to be able to model negative associations among clustered measurements. We present the *Bayesian Covariance Structure Modelling* (BCSM), a novel model that can deal with the issues of heterogeneity in individuals. The main advantage of BCSM is that the random effects are not included but implied dependencies modelled through the covariance matrix. By modeling dependencies between measurements and individuals directly, the BCSM can relax the assumption that random effect should a positive variance; in the BCSM, level two variances can also be negative.

## References


Fox, J. P., & Smink, W. A. C. (2019). *Bayesian Covariance Structure Modelling: A simulation study combined with real data to present novel method for multilevel modelling*. Oral presetation to be given at the 12th International Multilevel Conference, Utrecht, The Netherlands.

Lamers, S. M. A., Bohlmeijer, E. T., Korte, J., & Westerhof, G. J. (2015). The efficacy of life-review as online-guided self-help for adults: A randomized trial. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *70*(1), 24–34. https://doi.org/10.1093/geronb/gbu030

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*(2), 109–118. https://doi.org/10.1037/h0024436

Raudenbush, S. W., & Bryk, A. S. (2002). Applications in the Study of Individual Change. In *Hierarchical Linear Models* (2nd ed., pp. 160–204). Los Angeles, California: SAGE Publications Ltd.

# Sample size formulas for cluster randomized repeated measurement designs with p>2 levels

Teerenstra, S.[1*]

[1] Radboud university medical center, Radboud Institute for Health Science, Department for Health Evidence, group Biostatistics, the Netherlands

* Presenting author

**Suggested talk duration (min 15 max 20)**

**Summary (294 words)**

Cluster randomized repeated measurements designs like the pre-post, parallel group, or stepped wedge design can have a multi-level data structure, for example: patients in general practices. For 2-level structures like this, sample size formulas are available. However, multi-level data structures with more than 2 levels are found in practice. This has led, for example, to sample size formulas for cluster randomized (standard type) stepped wedge trials or cluster randomized parallel group design trials with 3 levels. We show a general approach to power and sample size calculation for cluster randomized repeated measurements designs using more than 2 levels.

By restricting to multilevel level structures build from random intercepts at levels, we show that the sample size and power can calculated using a so-called ‘design effect’ (variance inflation factor). This design effect is product of a design effect capturing the multilevel structure of the data, and a design effect ensuing from how intervention and control condition are allocated over time.

Apart from that, another key aspect is that there are different choices for which levels are measured as cohort and which cross-sectionally. For example in a 3-level cluster randomized trial with repeated measurements (evaluations within subjects within clusters), the same subjects could be repeatedly measured (cohort on level 2) or different ones (cross-sectional on level 2). We will see that this does not impact the generic form of the design factors, but is absorbed in the correlation over time of means from the same

cluster. This correlation in turn can be expressed by intraclass correlations. Thus, in the three level example above, the correlation of two evaluations within a subject and the correlation of two subjects within a cluster.

We will illustrate the impact of sample size and intraclass correlations at different levels on power in an example.

**Relevance to conference theme**

A substantial number of trials has a multilevel structure, for instance see the recent book by Mirjam Moerbeek and Steven Teerenstra. This abstracts bears on cluster trials with > 2 levels and repeated measurements. This is relevant across a range of disciplines ranging from education science to health care.

**Keywords (max. 3)**

Cluster randomized trial, repeated measures, power

# Missing data imputation in large combined cross-sectional and longitudinal data: multilevel multiple imputation and time series imputation

Wutchiett, David[1*]
[1] Université de Montréal, Canada
* Presenting author; PhD-student, advisor: Claire Durand

Suggested talk duration: 15-30 minutes

**Summary:**
Combining cross-sectional survey data and country-level data describing national socioeconomic conditions is frequently relevant for studies that evaluate both individual level and context related effects. However, missing values in both survey data and country-level data are very common in applied research. Where a multilevel model may be the model appropriate for statistical inference, differences in dependencies related to missingness in country-level and cross-sectional data introduce complications when imputing missing data. More specifically, with longitudinal data, imputed estimates using multivariate estimation are frequently non-optimal. Imputations may tend to be significantly closer to the overall mean values of the series than to temporally adjacent values in each country, and this despite inclusion of random effects for time variables. The data used is a large database of 550,000 cross-sectional survey respondents describing individual demographics and trust in democratic institutions in Eastern European and Eurasian countries across years 1991 to 2016. This database was combined with a database of yearly measures of each country's socioeconomic conditions. Following their combination, the process, methods and procedures for the imputation of yearly national summary statistics and survey respondent individual characteristics were evaluated. In regard to the imputation of national summary statistics, different specifications were considered, ranging from models accounting for only a country's unique longitudinal sequence of national observations to models incorporating multivariate multilevel specifications with random effects by country. Where time series imputation in many cases produced smooth and coherent imputations across sequential observations, larger spans of missing values tend to produce estimates that do not vary and differ much from the trends observed across other countries' time series. Tradeoffs concerning model specifications considering multiple countries or longitudinal trends through

the specification of fixed and random effects during an imputation procedure are further evaluated and discussed. Software involved will include R packages ‘mice’, ‘pan’, ‘imputeTS’, and ‘lme4’.

**Relevance to conference theme**

The research presentation will span the themes of innovative applications, software and methodology. Software involved are the “mice”, “pan”, “imputeTS” and “lme4” R packages. Methods will span multilevel multiple imputation for panel data and time series imputation. An original harmonized database of 550,000 cross-sectional survey responses and yearly panel data for national socioeconomic conditions will be used to apply the discussed methods and software.

**Summary:** Multilevel multiple imputation, Time series imputation, Longitudinal data, Cross-sectional data

# Testing replication of structural equation models

Zondervan-Zwijnenburg, M.A.J.[1*]

[1] Utrecht University, The Netherlands

* Presenting author (PhD student, Promotor: Herbert Hoijtink)

### Suggested talk duration (15-60 minutes)

25 minutes

### Summary (max. 500 words)

This presentation concerns testing replication of structural equation models (SEM), including multilevel models. Specifically, I will address what replication is, what current replication practices are, and how replication can be tested. That is, I propose to test the failure to replicate important findings of an original study by a new study with the prior predictive *p*-value. I will explain the steps that are taken in this procedure, and I will demonstrate how they can be easily executed in R with the Replication R-package. Furthermore, I will address the strengths and limitations of the prior predictive *p*-value with a multilevel latent growth curve model.

### Relevance to conference theme

The replication of structural equation models encompasses the replication of multilevel models. The main example that I will use is in fact a multilevel model (levels: time, parents, couples). Furthermore, I will present software (an R-package) with which researchers can apply the proposed method.

### Keywords (max. 3)

Replication, prior predictive *p*-value