# 8th International Amsterdam Conference Multilevel Analysis March 17 & 18, 2011

Scientific committee

Joop Hox, Mirjam Moerbeek, Rens van de Schoot & Leoniek Wijngaards

Local organisation committee

Kevin van Kats & Flip Boerland

Utrecht University
Department Methodology & Statistics

Financial support from the Royal Dutch Academy of Arts and Sciences (KNAW) is gratefully acknowledged

# MULTILEVEL CONFERENCE PROGRAMME

**Thursday, March 17**

| | |
|---|---|
| 9:00 | Registration |
| 9:25 | Opening |
| 9:30 | **Keynote presentation**: Stephen West<br>*Approaches to Multilevel Repeated Measures Data: Roles for Centering* |
| 10:10 | Eleni Sofianopoulou, et. al.<br>*Multilevel model with offset: A two stage approach for modeling respiratory prescribing, in Primary Health Care.* |
| 10:30 | Fiona Steele<br>*A multilevel simultaneous equations model for within-cluster dynamic effects, with an application to reciprocal relationships between maternal depression and child behaviour* |
| 10:50 | **BREAK** |
| | **PARALLEL SESSION I-a** |
| 11:10 | Omar Paccagnella, et. al.<br>*Profiling financial asset holders across Europe* |
| 11:30 | Adi Pierewan, et. al.<br>*Multivariate multilevel model of happiness and health across Europe during crisis* |
| 11:50 | Paul Clarke<br>*The choice between fixed and random effects multilevel models in policy-relevant social research* |
| 12:10 | Daniel Caro, et. al.<br>*A New Look at the Evaluation of Sociological Theories in International Large Scale Educational Assessments* |

**PARALLEL SESSION I-b**

11:10 Dickson Nkafu Anumendem, et. al.
*Comparison of Bivariate Multilevel Growth Curve Model and Bi-parallel Process Multilevel Latent Growth Model Approaches*

11:30 Sabine Meinck, et. al.
*Sampling Precision of Hierarchical Linear Models: Dependence on Varying Sample and Population Conditions*

11:50 Rok Platinovšek
*When is the accuracy worth the bother? A simulation-based comparison of two multiple imputation procedures applied to clustered datasets*

12:10 Hawjeng Chiou, et. al.
*Cluster mean replacement in the intercept and slope equations in the multilevel modeling: Simulation and empirical study of comparisons of the centering methods*

12:30 **LUNCH**

13:30 **Keynote presentation:** Bill Browne
*Statistical Software at the Centre for Multilevel Modelling*

14:10 George Leckie, et. al.
*Running MLwiN from within Stata: the* `runmlwin` *command*

14:30 Eva Batistatou, et. al.
*The performance of three methods of analysis in terms of power for trials with a partially nested design*

14:50 Chris Roberts
*The Analysis and Design of Non-Pharmacological Trials with Multiple-Membership due to Therapist*

15:10 **BREAK**

15:30 Neil H. Spencer
*Detecting a missing level in a multilevel model and assessing its impact*

15:50 Adam Steventon, et. al.
*The use of multilevel modelling to minimise likely hidden bias in the design of matched control studies*

16:10 Marianna Avetisyan
*Validation of the randomized response technique on multilevel data*

16:30 **END OF THE FIRST DAY**

**Friday, March 18**

8:30 **START OF THE CONFERENCE WITH COFFEE AND TEA**

9:00 Benjamin Kelcey
*Inferences on Instructional Practice Measured With Multiple Sources of Variation*

9:20 Ehri Ryu
*Factorial invariance in multilevel confirmatory factor analysis*

9:40 Steven Teerenstra et. al.
*A simple sample size formula for analysis of covariance in cluster randomized trials*

10:00 Esther de Hoop
*The simplest stepped wedge*

10:20 Gerard van Breukelen
*Efficient design of cluster randomized and multisite trials with unknown intraclass correlation and varying cluster size*

10:40 **BREAK**

**PARALLEL SESSION II-a**

11:00 Anna Cuxart
*Design Effect and Interviewer Effects: evidence from the Spanish ESS3-2006 Data*

11:20 Antonie Knigge, et. al.
*The total influence of family background on status attainment in the Netherlands from 1842-1922 – a multilevel sibling model*

11:40 George Leckie, et. al.
*The social relations model for dyadic count data, with an application to inter-household meat sharing in Nicaragua*

**PARALLEL SESSION II-b**

11:00 Victor H. Cervantes, et. al.
*Estimating item parameters for the dichotomous Rasch model with a complex sampling design*

11:20 Josine Verhagen
*A Bayesian Multilevel IRT Model for Testing Measurement Invariance in Cross-national and Longitudinal Surveys*

11:40 Margot Bennink, et. al.
*Micro-macro multilevel analysis for discrete data*

12:00 **LUNCH**

13:00 **Keynote presentation:** Douglas Bates
*Assessing precision of estimates through sensitivity analysis*

13:40 Elly Korendijk, et. al.
*Sample size requirements when applying a Fishbein-Ajzen Model in cluster randomized trials. A simulation study.*

14:00 Mark de Rooij
*The clustered trend vector model*

14:20 Ian Plewis
*Predicting Missingness in Longitudinal Studies: Can the Sample Design Help?*

14:40 **BREAK**

15:00 Tobias Koch, et. al.
*Structural Equation Models for Multitrait-Multimethod-Multioccasion (MTMM-MO) Data Combining Structurally Different and Interchangeable Methods*

15:20 Martin Schultze, et. al.
*Modeling longitudinal MTMM data with interchangeable and structurally different methods: An application to developmental psychology.*

15:40 Elif Ünal Çoker, et. al.
*Exploring the Performance of Information Criteria in Multilevel Structural Equation Modelling*

16:00 **THE END OF THE MULTILEVEL CONFERENCE 2011**

# Comparison of Bivariate Multilevel Growth Curve Model and Bi-parallel Process Multilevel Latent Growth Model Approaches.

Anumendem, Dickson Nkafu, Bieke De Fraine,
Patrick Onghena and Jan Van Damme

There is an increasing interest in the field of educational effectiveness research for studying changes in pupils' outcomes. However, most studies focus on the growth in only one effectiveness criterion, which is problematic given that school effects are only moderately consistent over different criteria. Multiple criteria have seldom been studied through multivariate growth curve models. The current study investigates school effects on pupils' growth in mathematics and reading comprehension in primary schools through the use of the two contemporary and competing frameworks for the analysis of growth curve model. The bi-parallel process multilevel latent growth model which models two dependent outcome variables as multivariate variable with the repeated measurements of the two variables all becoming variables in themselves and all assumed independent. This is unlike the bivariate multilevel growth curve model which considers independence only for the repeated measurements of the same outcome variable but allows the dependence of the two growth processes. The later enables a statistical test for the need of a multivariate growth model over separate univariate growth models. Of course constraints can be used in both approaches but then deviations from the fundamental theory behind them are inevitable. By using the proposed two-stage effectiveness criteria, we found some consistency between school effects criteria. However, the comparison between the two approaches reveal that most of the similarities found in univariate repeated measurements growth curves, do not simply export themselves into the multivariate repeated measurements in terms of their theoretical framework as well as applications to real data.

Keywords: Educational effectiveness, effectiveness criteria, reading comprehension, mathematics, growth curve models, multilevel analysis, bivariate multilevel growth curve model, bi-parallel process multilevel latent growth model.

---

Dickson.nkafu@ped.kuleuven.be

[1]*Centre for Educational Effectiveness and Evaluation, K.U.Leuven, Belgium*

[2]*Centre for Methodology in Educational Sciences, K.U.Leuven, Belgium*

# Validation of the randomized response technique on multilevel data

**Avetisyan, Marianna**

Survey respondents usually tend to conceal embarrassing or threatening information, particularly when an interviewer might disapprove. This response behavior is certainly expected in medical settings where the incriminating behavior is negatively related to the recovery from a disease. The lung patient group chosen for our research feels social (public smoking policy) and professional (pulmonologists) pressure concerning nonsmoking and therefore can experience questions about smoking behavior as sensitive. For this study a questionnaire measuring attitude towards smoking was developed. The primary goal of this research was the validation of the randomized response technique, a method of administration of a questionnaire stimulating honest responses to sensitive questions. For validation purposes in the first stage of the research the subjects filled in the questionnaire in a conventional way, i.e. by direct questioning, or using the randomized response technique followed by the second stage when a breath test using a carbon monoxide (CO) monitor was taken revealing the actual smoking status of the subject. The observed responses are hierarchical in nature with the three-level hierarchy of measurements nested within patients nested within pulmonologists. These hierarchical data are analyzed using IRT measuring the attitude toward smoking.

Key words: randomized response, item response theory, multilevel modeling, Bayesian inference.

---

*Marianna Avetisyan*
*University of Twente*
*Faculty of Behavioural Sciences*
*Department of Measurement and Data Analysis*
*P.O. Box 217, 7500 AE Enschede, The Netherlands*
*m.avetisyan@gw.utwente.nl*

# Assessing precision of estimates through sensitivity analysis

## Bates, Douglas

Fitting a mixed-effects model to observed data provides point estimates of the parameters. As part of the statistical analysis we should also describe the precision of the parameter estimates, say through confidence intervals, and provide methods for performing hypothesis tests. Frequently confidence intervals on coefficients and on variance-component parameters are constructed as symmetric intervals derived from an approximate standard error and the quantiles of a t or z distribution. There is some theoretical justfication for such intervals on the coefficients of models fit to balanced data but the arguments fail in the case of coefficients for models fit to unbalanced, observational data and in all cases for variance-component parameters. Even in the simplest case of estimating the variance of a normal distribution based on a random sample we realize that confidence intervals will be highly asymmetric. Expecting that the variability in estimates of variance components for much more complex models will be adequately summarized by a standard error leading to symmetric intervals is, at best, optimistic. We present methods of assessing the variability by profiling the likelihood, which is a form of "sensitivity analysis". That is, we examine the sensitivity of the estimation criterion - the deviance, in this case - to changes in individual parameters. The deviance profiles allow for assessment of the precision of individual parameters and for projections of contours of the deviance on pairs of parameters.

# The performance of three methods of analysis in terms of power for trials with a partially nested design

**Batistatou, Eva; Roberts, Chris**

Partially nested data, where forming clusters of individuals for the purpose of the clinical trial in pthe experimental condition only, arise in trials of group administered treatments. It is well known that application of conventional sample size estimation formulae which do not account for the cluster size and the intraclass correlation coefficient (ICC) within experimental condition may lead to sample size estimates that are too small and trials that are inadequately powered. Additionally, the outcome variance in the two trial arms may not be equal, leading often to substantial heteroscedasticity between arms. Calculating power by integrating the noncentral F-distribution1 – implemented in our clsampsi Stata's program2 – may be more appropriate. Here we report results from a simulation study for partially nested continuous data that evaluated the performance of three analysis approaches in terms of empirical power: the partially nested model via REML and ML estimation methods and the unequal variance t-test applied to a cluster-level summary; various numbers of clusters, cluster sizes, ICCs for the intervention arm and trial arm variance ratios were used. The empirical power calculated via simulations was compared with that calculated through our clsampsi program. It was shown that using the t-test for analyzing partially nested data attains empirical power very close to that calculated by clsampsi for all the parameter scenarios tested, while the other two approaches of analysis give either larger or considerably larger empirical power for almost all the scenarios.

[1] Roberts & Roberts (2005).

[2] http://www.medicine.manchester.ac.uk/healthmethodology/research/biostatistics/data/clsampsi/

Key words: partially nested data, multilevel models, group therapy

---

*Dr Eva Batistatou - Research Associate - Health Sciences Research Group School of Community Based Medicine - University of Manchester, UK - E.Batistatou@manchester.ac.uk*

*Dr Chris Roberts - Reader in Biostatistics - Health Sciences Research Group School of Community Based Medicine - University of Manchester, UK – Chris.Roberts@manchester.ac.uk*

# Micro-macro multilevel analysis for discrete data

**Bennink, Margot; Croon, Marcel A.; Vermunt, Jeroen K.**

This study deals with models for predicting outcomes at the higher level (e.g. team performance) from explanatory variables at the lower level (e.g. employee's motivation and skills). This "reversed" multilevel analysis problem is rather common in social sciences, and is sometimes referred to as micro-macro analysis. Recently, Croon and Van Veldhoven proposed a statistical model for micro-macro multilevel analysis which involves using a factor analytic structure in which the scores of the lower-level units are seen as indicators of latent factors at the group level. The key is that the outcome variable is not regressed on the aggregated group mean(s) of the micro-level predictor(s) but on the latent macro-level variable(s). The aim of the project, from which the current study is a part, is to generalize this approach so that it can also be applied when the explanatory and/or outcome variables are discrete instead of continuous and normally distributed. Two new models for micro-macro relations between discrete variables are presented; a simple 1-2 model in which a dichotomous micro-level variable affects a dichotomous macro-level outcome variable, and a more complex 2-1-2 model in which a dichotomous macro-level variable has a direct effect on a dichotomous macro-level outcome variable and an indirect effect on the outcome through a dichotomous mediating variable defined at the micro-level. In both models the latent variable at the group level is defined to be discrete (latent classes). We present the theoretical background of the models, a simulation study in which their performance is evaluated, as well as an empirical application.

Key words: multilevel analysis, micro-macro relationships, dichotomous outcome variables, discrete data

---

*Margot Bennink*
*Phone: +31 13 466 8046*
*Secretary: +31 13 466 2544*
*m.bennink@uvt.nl*

*Tilburg University*
*Faculty of Social and Behavioural Sciences*
*Department Methodology and Statistics*

*Room P1121*
*PO Box 90153*
*5000 LE Tilburg*

# Efficient design of cluster randomized and multisite trials with unknown intraclass correlation and varying cluster size

**Breukelen, Gerard van**

Objective:
Optimal sample size equations for cluster randomized trials (how many clusters, how many persons) assume equal cluster sizes and a known intraclass correlation (*ICC*). A simple and accurate adjustment for varying cluster sizes will be demonstrated. The assumption of a known *ICC* is problematic since optimal design depends on this *ICC* and is thus a locally optimal design (LOD). We present and compare simple solutions to this problem.
Method:
A closed form is presented for the Maximin design (MMD) given an unknown *ICC* with user-specified range [*a, b*], for instance $a = 0.01$ and $b = 0.10$. The MMD maximizes the minimum relative efficiency compared with the LOD, where the minimum is taken over the whole *ICC* range and corresponding set of LODs. The MMD is compared with some simple alternatives, such as the LOD for $ICC = b$ (called LOD(*b*)), with respect to sample sizes and performance.
Result:
The minimum relative efficiency of the MMD decreases as the *ICC* range increases, but is always much higher than for the LOD(*b*), while the latter has a higher minimum efficiency. A simple alternative in-between MMD and LOD(*b*) performs fairly well, whereas the LOD(*a*) performs badly, on efficiency and relative efficiency.
Conclusion:
Maximin design improves robustness against misspecified *ICC* at the price of some loss of efficiency compared with the LOD for the correct but unknown *ICC*. By choosing an *ICC* range, MMD allows balancing between efficiency and robustness.

Keywords: cluster randomized trials, optimal design, Maximin design, intraclass correlation, relative efficiency

---

*Address:*
*Gerard van Breukelen, Dept. Methodology & Statistics, Faculty of Psychology & Neuroscience, and Caphri School for Public Health and PrimaryCcare of the Faculty of Health, Medicine and Life Sciences, Maastricht University, p.o. box 616, 6200 MD Maastricht, The Netherlands*

# Statistical Software at the Centre for Multilevel Modelling

**Browne, Bill**

In this talk, a general history of developments in statistical software at the Centre for Multilevel Modelling, first at the IOE in London and then continued in Bristol will be given. We will discuss developments that lead to the MLwiN package and other projects like the MLPowSim and REALCOM packages. We will then describe our ambitious new ESRC funded E-STAT project which builds on initial concepts from Jon Rasbash and is a collaboration with academics from Southampton and several other UK universities. We will describe how the project aims to inter-operate with other software packages and demonstrate the current incarnation of the software.

# A New Look at the Evaluation of Sociological Theories in International Large Scale Educational Assessments

**Caro, Daniel***
**Sandoval-Hernandez, Andres***

The paper uses a mixed approach of theory generating and theory testing to evaluate theoretical models of cultural and social capital of Bourdieu, Bernstein and Coleman. The data stem from two large scale international assessments: PIRLS 2006 and PISA 2009. In line with the mixed approach employed, the method is a combination of exploratory factor analysis (EFA) and structural equation models (SEM) known as exploratory structural equation modeling (ESEM). The main advantage of using this method is that an EFA measurement model with rotations can be used in a SEM, instead of a confirmatory factor analysis (CFA) measurement model. When using the EFA we have the advantage of being able to relax the zero loading restriction, allowing variables to load on different factors simultaneously. The SEM then provides model fit indices, yields measurement invariance tests across countries, and corrects factor scores for measurement error. The results provide statistical criteria to evaluate the fit of theories with the empirical data (i.e., more reliably and validly). The paper contributes to the discussion on the advantages and disadvantages of using these methods when evaluating sociological theories in the context of international large scale assessments.

**Keywords**: exploratory structural equation modeling, cultural capital, cross-cultural validity, PIRLS

---

** International Association for the Evaluation of Educational Achievement (IEA), Data Processing and Research Center (DPC), Hamburg, Germany*.

# Estimating item parameters for the dichotomous Rasch model with a complex sampling design

**Cervantes, Víctor H. and Uzaheta, Álvaro**

Item response models (IRM) may be seen as special cases of multilevel models often used in psychological and educational assessment. In educational assessment they are usually used with complex sampling designs where students are themselves clustered within schools, these within cities and countries, in a hierarchical structure. Among IRM, Rasch models are used in many international and national assessments such as PISA and SERCE. Some of these assessments estimate item parameters from sub-samples without using the sampling weights associated with the original sampling design, nor the ones associated with the further second stage sampling employed. While some concern on the proper use of weights in this context has been posed (e.g. Cyr & Davies, 2005), we have not found in the literature general prescriptions on how to incorporate such weights. We seek to assess the precision of items' estimates under different weighting procedures with a complex sampling design and different estimation approaches for the Rasch model. A Monte Carlo simulation study where a complex sampling design similar to the one used in Colombia's 2009 Saber 5º y 9º will be used to assess the procedures. We expect the results from this study to shade light into which weighting scheme may be more appropriate to incorporate the expansion weights produced by the sampling design into the item parameter estimation for future Colombian educational assessments as well as for other studies, such as those mentioned.

Cyr, A. & Davies, A. Paper presented in the 2005 FCSM Conference.

Key words: item response models, complex sampling, sampling weights, Rasch model, Winsteps, lme4

---

*Addresses. vcervantes@icfes.gov.co, auzaheta@icfes.gov.co. Instituto Colombiano para la Evaluación de la Educación – ICFES- Colombia*

# Cluster mean replacement in the intercept and slope equations in the multilevel modeling: Simulation and empirical study of comparisons of the centering methods

**Chiou, Hawjeng** ; **Lin, Pi-Fang**,

In multilevel linear modeling (MLM), the issue of centering Level-1 predictor variables is critical. In literature, it has been argued that both of centering methods, centering at the grand mean (CGM) and centering within cluster (CWC) has advantages and disadvantages. In short, CWC is preferable if the effect of Level 1 predictor of X alone on a Y as outcome, and if the cross-level interactions and level-1 interactions are of substantive interest. As a result, the cluster mean of X should be replaced back into the level-2 as a main effect in the intercept equation ($\beta_{0j}$) because CWC removes all between-cluster variation in X. However, the present paper argued that the replacement of cluster mean in intercept equation is insufficient. Because it may exist a significant cross-level interactions of X and cluster mean of X, which can only be detected by an interaction term of X and cluster mean of X. In other words, replacement of cluster mean of X into the slope equation ($\beta_{1j}$) along with intercept equation has to be created for sufficiently examining the effect of CWC.

In order to compare with preview study, we used the artificial data set with 300 clusters of 40 cases each created by Enders and Tofighi (2007) as a dataset 1. Results of a mixed model using dataset 1 were comparing with dataset 2 in which a special treatment is made for creating a strong interaction term of X and cluster mean of X. Results revealed that, in dataset 1, there was only main effect of the cluster mean of X on Y. No significant interaction effect of the X and the cluster mean of X in dataset 1 was reported due to the small effect. In dataset 2, an artificial interaction between X and the cluster mean of X was created and, as a result, the interaction term was significant. That is, the replacement of the cluster mean into the slop equation is necessary to detect the effect of interaction of deviation score and its own mean. Besides of the evidences of simulation data analysis, an empirical data selected from author's previous study consisted of 664 employees sampled from 24 different organizations were applied. Results revealed that the CWC with replacement of the cluster mean into the intercept and slope equation at macro-level was the best estimates of fixed effects as well as random effects. Implications and limitations of the study were discussed in the final section.

Key words: group-mean centering, grand-mean centering, cross-level interaction, replacement of the cluster mean

*Author information:*

*Hawjeng Chiou, Ph.D.*
*Associate professor*
*College of Management*
*National Taiwan Normal University, Taiwan*
*Email: hawjeng@ntnu.edu.tw*

*Pi-Fang Lin, Doctoral candidate*
*Graduate School of Education*
*National Cheng Chi University, Taiwan*
*Email: 95152518@nccu.edu.tw*

# The choice between fixed and random effects multilevel models in policy-relevant social research

**Clarke, Paul**

The use of fixed and random effects multilevel models is discussed in the context of policy-relevant social research, and the assumptions behind the two approaches set out. By explicating the key issues we hope to resolve some of the confusion surrounding the differences between fixed and random effects multilevel models. To illustrate these issues, an example from education - analysing the determinants of pupil achievement in primary school - is used. It is concluded that a fixed effects approach will be preferable in scenarios where the primary interest is in policy-relevant inferences about the effects of individual characteristics, and the process through which individuals are selected into clusters (in this case schools) is either poorly understood or the data too limited to adjust for selection effects. In this context, the robustness of the fixed effects approach to the random effects assumption is attractive, and researchers should consider using it if only to assess the sensitivity of the random effects estimates to assumption failures. On the other hand, when the selection mechanism is better understood and the researcher has access to rich data, a random effects approach is more natural because it can be used to produce policy-relevant estimates while allowing a wider range of research questions to be addressed; moreover, random effects estimators of regression coefficients and shrinkage estimators of cluster effects are more efficient than those for fixed effects.

Key words: Causal inference; Confounding bias; Fixed effects; Random effects; Random effects assumption; Robustness

---

*Presenting author: Paul Clarke, Centre for Market & Public Organisation, University of Bristol, UK*
*Co-authors: Clarie Crawford, Institute for Fiscal Studies, UK; Fiona Steele, University of Bristol, UK; Anna Vignoles, Institute of Education, University of London, UK.*

# Design Effect and Interviewer Effects: Evidence from the Spanish ESS3-2006 Data

**Cuxart, Anna**

In 2007 the first Quality Enhancement Meeting on sampling in the European Social Survey (ESS) took place. The discussion focused on design effects and inteviewer effects in face-to-face interviews. Following the recomendations of this meeting the Spanish ESS team studied the impact of interviewers as a new element in the design effect in the response's variance using the information of the correspondent Sample Design Data Files. Hierarchical multilevel and cross-classified multilevel analysis are conducted in order to estimate the amount of responses' variation due to PSU and to interviewers for different questions in the survey. Factor such as the age of the interviewer, gender, workload, training and experience and respondent characteristics such as age, gender, renuance to participate and their possible interactions are also included in the analysis of some specific questions like "trust in politicians" and "trust in legal system". Some recomendations related to future sampling designs and the contents of the briefing sessions are derived from this initial research.

Key words: interviewer effects, multilevel analysis, design effect, European social survey

---

*Addresses, sites etc.*
*anna.cuxart@upf.edu*

*Universitat Pompeu Fabra,*
*Economica and Business department,*

*Campus de la Ciutadella,*
*Ramon Trias Fargas 25,*
*08005 Barcelona,*
*Spain*

# The simplest stepped wedge


## Hoop, Esther de



Cluster randomized trials are known to be less efficient than individually randomized trials, especially if only the outcomes at the end of the study are analyzed (posttest design). Using covariates or repeated measurements can improve the efficiency. In practice, this is feasible, because often a baseline measurement of the outcome is available or easy to obtain, as well as an additional follow-up score.
We compare three designs that use repeated measurements or covariates with the posttest design: an ancova model with baseline measurement as covariate, an extension of this ancova model with a follow-up measurement, and a simple stepped wedge model.
To compare the efficiency of these designs, we derived variance inflation factors for the three extended models in terms of $r$, which is the correlation of the cluster means between baseline and follow-up.
The variance of an ancova analysis is a factor $r^2$ smaller than of an analysis on follow-up scores only. A stepped wedge design results in a small gain in efficiency compared to the ancova model. The ancova model with an additional follow-up measurement is much more efficient than all other methods. As the correlation increases, all models with more than one measurement become more and more efficient. Since in many cases correlation between baseline and follow-up is expected, it is worthwhile to consider research designs with additional measurements.



Key words: cluster randomized trials, efficiency, repeated measurements, ancova, stepped wedge


---


*Radboud University Nijmegen Medical Centre*
*Department of Epidemiology, Biostatistics and HTA 133*
*PO Box 9101*
*6500 HB Nijmegen*
*The Netherlands*
*E-mail: E.deHoop@ebh.umcn.nl*

# Inferences on Instructional Practice Measured With Multiple Sources of Variation

**Kelcey, Benjamin**

Recent educational research has developed a number of observation and coding systems to chronicle enacted teaching. Such systems detail enacted teaching within each lesson by having trained observers record the presence/absence of targeted behavioral features thought to reflect central dimensions of teaching. The expectation is that such systems will help draw inferences as to the nature of instruction across a year using a sample of lessons and days to uncover reliable evidence concerning the processes which drive teachers' contribution to students' growth. To form indices of instruction from such systems, we outline the conceptualization and application of a framework that uses a cross-classified mixed effects measurement model. We use item response theory and generalizability theory to describe how the framework draws inferences from the sample of lessons and then show how it tends to produce more reliable indices of practice than alternative approaches when faced with nonrandom measurement error (e.g. rater or observation day bias). In turn, we link this increased reliability with increased power to detect contextual effects on enacted practice and then identify avenues by which we might improve upon reliability in subsequent studies (e.g. adding raters). To make details more concrete, we consider an application of the framework to the study of incidental word instruction in early elementary school. Here, trained observers recorded teachers' use of targeted instructional actions (thought to be reflective of one aspect of strong literacy instruction) across multiple lessons and multiple days throughout the school year.

Key words: Cross-classified mixed models, measurement of instruction, g-theory, item response theory, ecometrics

---

*Addresses, sites etc. Wayne State University, 5425 Gullen Mall, #347 COE, Detroit, Michigan, USA 48202; ben.kelcey@gmail.com*

# The total influence of family background on status attainment in the Netherlands from 1842-1922 – a multilevel sibling model

**Knigge, Antonie, Ineke Maas, Marco van Leeuwen & Kees Mandemakers**

The modernization thesis claims that with industrialization and other modernization processes the influence of the family on the status attainment of their children has decreased. In this study we use a new approach to test whether society indeed has become more open over time for the case of the Netherlands, and whether such changes in the family influence can be explained by modernization processes. The first major contribution of this paper is that we use multilevel sibling models to assess the family influence on status attainment. This method allows us to measure the *total* family influence, whereas using the conventional indicator (occupational status of the father) will underestimate the true influence. The second major contribution of this paper is that we use a large-scale historical database that contains information from Dutch marriage records of approximately 500 000 sons from 300 000 families while covering more than 500 municipalities and an 80 year period (1842-1922). We complement these individual and family level data with regional indicators for industrialization, mass communication, urbanization, geographical mobility and mass transport. Preliminary results seem to indicate that the family influence in Dutch society increased at a very low rate until around 1870, and then started to decrease slowly but steadily.

# Structural Equation Models for Multitrait-Multimethod-Multioccasion (MTMM-MO) Data Combining Structurally Different and Interchangeable Methods

**Koch, Tobias[1]; Schultze, Martin[1]; Geiser, Christian[2]; Eid, Michael[1]**

The Multitrait-Multimethod (MTMM) approach (Campbell & Fiske, 1959) is one of the most vital methodological developments in validation research. Currently, the most common way to analyze MTMM data is via structural equation models (SEMs). Eid, Nussbeck, Geiser, Cole, Gollwitzer, and Lischetzke (2008) recently clarified that different measurement designs imply different statistical models and discussed appropriate SEM-MTMM models for cross-sectional designs with different types of methods. In particular, measurement designs with *interchangeable* methods use methods that are randomly chosen out of a set of equivalent methods (e.g., multiple student ratings of teaching quality). Hence, measurement designs with interchangeable methods imply a multilevel data structure. In contrast, measurement designs with *structurally different* methods use fixed methods, meaning that one method cannot easily be replaced by one of the other methods. Accordingly, physiological measures, self-ratings, and teacher ratings can be considered as structurally different methods. An increasing number of psychological studies use a combination of interchangeable and structurally different methods within a longitudinal research design. We present a new SEM-MTMM model for analyzing longitudinal MTMM data with interchangeable and structurally different methods. In this talk, we want to clarify the basic idea of the model and explain the meaning of consistency, method specificity, and latent correlations coefficients. The advantages as well as the limitations of the model will be discussed.

Keywords: MTMM-MO, structural equation modeling, multilevel modeling, interchangeable methods, structurally different methods

---

***[1] Freie Universität Berlin; [2] Arizona State University.***

# Sample size requirements when applying a Fishbein-Ajzen Model in cluster randomized trials. A simulation study.

**Korendijk Elly[1], Joop Hox[2] & Mirjam Moerbeek[2]**

In experimental research the Fishbein-Ajzen model is used to design and evaluate interventions. The model states that behavior of individuals is affected by the intention of subjects, which in its turn is affected by the attitude toward the behavior and the perceived social norm of the individual. In experimental research it is thought that the intervention affects the intention via the attitude, i.e. attitude mediates the affect of intervention on intention. Since intention mediates the effect of attitude and norm on behavior, the model now consists of two mediator variables in sequence. In cluster randomized trials the intervention is a cluster level variable, while the predictor variables in the Fishbein-Ajzen model, i.e. attitude, perceived social norm and intention, as well as the outcome variable behavior, are subject level variables. The multilevel model at hand with two mediators in sequence and the intervention at the second level, is a rather complex model and needs a reasonable sample size in order to detect the intervention effect. Research, focusing on power and sample size in the frame work of multilevel modeling, showed that the sample size at the cluster level is more important than the cluster size. The Fishbein-Ajzen model and its effectiveness is described in a substantive amount of publications since the introduction of the model in 1975 and hence also in the frame work of multilevel modeling. However, so far it is not investigated which sample sizes are required to establish a mediated intervention effect for the Fishbein-Ajzen model in the frame work of a cluster randomized trial, i.e. when the intervention is at the subject level. In the current study, this issue is addressed by means of a simulation study.

---

[1] Corresponding author. Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University. P.O. Box 80140, 3508 TC Utrecht, the Netherlands, e.j.h.korendijk@uu.nl

[2] Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, the Netherlands.

# The social relations model for dyadic count data, with an application to inter-household meat sharing in Nicaragua

**Leckie, George; Jeremy Koster**

The Social Relations Model (SRM) has proved a popular approach to study dyadic data on the relationships between pairs of individuals. Specifically, the SRM decomposes the relationship response variance into sender-, receiver- and unique relationship-specific effects. Additionally, the approach models both the correlation between the sender and receiver effects and the correlation of responses within dyads.

Surprisingly, applications of the SRM have been almost universally limited to the analysis of continuous responses despite relational or dyadic data often being discrete. In this presentation, we shall describe a multilevel formulation of the SRM for count data and discuss the interesting methodological features of the resulting model: it is a constrained bivariate poisson response model for cross-classified data with correlated random effects across classifications.

We shall illustrate the SRM model for count data with an application to inter-household meat sharing in Nicaragua. Specifically, we model the counts of meat shares between 35 households from two neighbouring indigenous communities over a nine month period. We relate these counts to, among other variables: households' community of residence, household size and composition and the genetic relatedness of households.

Key words: social relations model, cross-classified model, correlated random classifications, MCMC, count data, poisson multilevel model

---

*George Leckie*
*Centre for Multilevel Modelling*
*University of Bristol*
*2 Priory Road*
*Bristol, BS8 1TX*
*United Kingdom*
*g.leckie@bristol.ac.uk*

*Jeremy Koster*
*Department of Anthropology*
*University of Cincinnati*
*Cincinnati, OH 45221-0380*
*United States*
*jeremy.koster@uc.edu*

# Running MLwiN from within Stata: the runmlwin command


## Leckie, George; Chris Charlton


The Centre for Multilevel Modelling is developing `runmlwin`, a Stata command to fit multilevel models in MLwiN from within Stata. There are three steps to using `runmlwin`: (1) The researcher specifies the desired model using the `runmlwin` command syntax; (2) The model is sent to and fitted in MLwiN; and (3) The results are returned to and displayed in Stata where they can be accessed for further analyses.

`runmlwin` will benefit Stata users by enabling them to fit a considerably wider range of multilevel models than they can currently and to fit these models quickly and to large data sets using fast estimation engines. Stata users can then examine these models using the many interactive tools available in MLwiN.

`runmlwin` will also benefit MLwiN users familiar with Stata as they can now type all the commands for their analysis into a single file and to run them all at once. This makes it easy to document and reproduce the results for an entire series of MLwiN models. MLwiN users can then make use of Stata's many inbuilt post-estimation commands to calculate predictions, perform hypothesis tests, and produce publication quality graphics. Even simulation studies are now easy to perform.

In this talk, we shall provide an overview of the `runmlwin` command and then demonstrate `runmlwin` in action with several example multilevel analyses.


Key words: MLwiN, multilevel model, `runmlwin`, Stata


---


*Centre for Multilevel Modelling*
*University of Bristol*
*2 Priory Road*
*Bristol BS8 1TX*
*UK*

*g.leckie@bristol.ac.uk*

# Sampling Precision of Hierarchical Linear Models: Dependence on Varying Sample and Population Conditions

**Meinck, Sabine & Vandenplas, Caroline**

This study explored the relationships between sample sizes and sampling precision of hierarchical linear model parameters, focusing on the properties of data collected in large-scale educational surveys. A Monte Carlo simulation study was used in order to explore various population and sample conditions. These conditions were varied regarding sample sizes of and within clusters, intraclass correlation, covariance distribution, use of sampling weights and model complexity.
The precision of all explored parameters increased with increasing sample sizes in non-linear format. This general observation held true for all settings. The magnitude of the increase, and whether the effect is more pronounced with sample size increases on either of the hierarchical levels, can depend however on all explored sample and population conditions and varies for the different model parameters. In conclusion, the results show that required sample sizes depend heavily on the parameter of interest. The effect of the covariance distribution on the sampling precision of certain model parameters appears to be even more pronounced than the effect of varying sample sizes. The inclusion of sampling weights in the model decreases the sampling precision of all explored parameters by approximately 10%. The model complexity had an influence on the sampling precision of all observed parameters except of the residual variance. This influence varied with the parameter of interest as well as with the considered case of covariance distribution.
All findings are only valid within the explored ranges of sampling and population settings. The degree of their generalizability shall be subject of further research.

*HLM, Variance Estimation, Sample Size, Intraclass Correlation, Sampling Weights, Large-Scale Assessments*

---

International Association for the Evaluation of Educational Achievement, Data Processing and Research Center (IEA DPC), Hamburg, Germany

# Profiling financial asset holders across Europe

**Paccagnella, Omar and Varriale, Roberta**

Wealth is an useful measure of the socio-economic status of the elderly, because it might reflect both accumulated socio-economic position and potential for current consumption. Even tough the elderly control, as a group, a substantial fraction of the household wealth, the majority of them has restricted portfolio holdings and does not invest in risky financial assets. Furthermore, there is evidence of a large heterogeneity in household portfolio ownership within and across European countries. The segmentation of the European market has been widely investigated over the last decade: the identification of homogenous segments helps companies to develop and implement international marketing strategies and the findings can support the researches on the household life-cycle effects within the financial market.

In this paper, we aim at providing new evidence on this issue both at the household and country level, by investigating similarities and differences in the ownership patterns of several financial assets among elderly in Europe. To do so, we exploit the richness of information provided by SHARE (Survey of Health, Ageing and Retirement in Europe), an international survey on ageing collecting detailed information on several aspects of the socioeconomic condition of the European elderly.

Given the hierarchical structure of the data, the econometric solution we adopt is a multilevel latent class analysis, which allows obtaining country and household segments simultaneously. In particular, the role of the households' characteristics in their choices of asset ownership is investigated. Models with and without first-level covariates are estimated and ex-post household profiles are evaluated and compared.

Key words: Ageing, household profiles, latent class analysis, multilevel data analysis, segmentation

---

*Omar Paccagnella, Department of Statistical Sciences, University of Padua, via C. Battisti, 241/243 - 35121 Padova, Italy, e-mail: omar.paccagnella@unipd.it*
*Roberta Varriale, Department of Statstics "G.Parenti", University of Florence, Viale Morgagni, 59 - 50134 Firenze, Italy e-mail: roberta.varriale@ds.unifi.it*

# Multivariate multilevel model of happiness and health across Europe during crisis

**Pierewan, Adi and Gindo Tampubolon**

Most studies on happiness and health tend to separate them as a single outcome. This study investigates the individual and the country level covariation in happiness and health simultaneously as outcomes. Using data from the 2008 European Values Survey, including 56210 respondents and 39 countries, we use multivariate multilevel model. We find that both at individual and country levels, happiness and health are positively correlated both at individual and country levels. We also find that most determinants of both happiness and health are relatively similar. We also find that in individual level: educated, married, and higher income people are positively associated with happiness and health. Conversely, unemployed and older people are negatively associated with happiness and health. In country level, we find that GDP and GDP growth are weakly and positively related with happiness and health; and Eastern Europe countries are weakly and negatively related with happiness and health.

Keywords: multivariate multilevel, happiness, health, well-being, crisis

---

Addresses:
Institute for Social Change, The University of Manchester,
Oxford Road M13 9PL, United Kingdom
Email: adi.pierewan@postgrad.manchester.ac.uk
Website: http://www.humanities.manchester.ac.uk/socialchange/

# When is the accuracy worth the bother? A simulation-based comparison of two multiple imputation procedures applied to clustered datasets

**Platinovšek, Rok**

In recent years, procedures have become available that perform multiple imputation (MI) of missing values under a multilevel model, thus allowing MI of clustered datasets while taking the appropriate variance structure into account. The methods in question, however, are computationally intensive and require a fair degree of statistical and programming sophistication on the part of the user. A recent paper has demonstrated that in certain cases a much simpler MI procedure can be used with negligible consequences for the resulting inference.

The presented study uses simulated data to compare the inferences drawn after imputing the same dataset with two competing MI procedures: one that takes the multilevel structure of the data into account and one that ignores it. The simulation study varies three factors: i) the number of clusters and units, ii) the intra-cluster correlation, and iii) the proportion of missing values. The quantity of interest is the beta-coefficient in a varying-intercept model that is fit to each MI dataset.

The results show that under ordinary conditions of moderate intra-cluster correlation and moderate proportion of missing values, the simpler MI procedure performs as well as its multilevel counterpart. Only when a substantial proportion of the data is missing *and* the intra-cluster correlation is high, does the coverage rate (CR) of the simpler procedure's confidence intervals drop below the nominal. The results are summed-up graphically by tracing the border-line where the CR for the simpler procedure drops below 0.90.

Key words: Multiple imputation, Multilevel models, Missing covariate, Coverage rate, Confidence interval

*rok.platinovsek@fdv.uni-lj.si*
*University of Ljubljana, Slovenia*

# Predicting Missingness in Longitudinal Studies: Can the Sample Design Help?

**Plewis, Ian**

Links between the analysis of clustered survey data and multilevel modelling are well-established. In essence, the survey tradition treats the clusters as nuisances to be allowed for in order to estimate standard errors correctly (by, for example, using the *svy* commands in STATA) whereas multilevel modellers recognise that the clusters can carry information that should be included in the model of interest. The same kinds of arguments can be used when modelling the propensities of not responding in any kind of survey: either allowing for the sample design in logistic or probit models for non-response or modelling the cluster effects in two level models. Longitudinal studies offer ways of modelling non-response not available in cross-sectional designs because information from earlier waves can be used to predict different kinds of non-response in the current wave. In these circumstances, can information about the clusters sampled at wave one add anything to these predictions? The paper will provide an answer to this question by drawing on data from the first two waves of the UK Millennium Cohort Study to assess predictiveness in terms of areas under Receiver Operating Characteristic (ROC) curves generated from predicted probabilities of not responding from two level multinomial logistic models, with the effects for each cluster explicitly estimated using MCMC in MLwiN. This work emanates from a project funded by the UK Economic and Social Research Council as part of its Survey Design and Measurement Initiative.

Keywords: cluster effects, longitudinal studies, multinomial models, non-response

---

*ian.plewis@manchester.ac.uk*

Social Statistics
University of Manchester, UK

# The Analysis and Design of Non-Pharmacological Trials with Multiple-Membership due to Therapist

**Roberts, Chris**

In trials of physical and talking therapies, nesting of patients within care-provider has design and analysis implications analogous to cluster randomized trials. Where patients receive treatment from a single care-provider, statistical analysis can be based on hierarchical models. In some trials, a patient may receives a therapy from multiple care-providers, so that the data structure is no longer hierarchical. In such a design, treatment effects can be estimated using multiple-membership models. An alternative is a primary therapist analysis carried out using the nesting by the patient's main care-provider.

These two approaches are compared in a simulation study of partially nested trial designs for moderate and large effect sizes. High rates of convergence rates can be achieved. As with simpler random intercept models estimates of the intra-cluster correlation coefficient (ICC) for care-provider are biased close to zero. Together with uncertainty in the between care-provider variance this can lead to either positive or negative biases in test size, power and coverage, but these tend to be smaller than for primary therapist analyses.

Whilst multiple-membership models give better estimate of treatment than a primary therapist analysis, there are design considerations as the analyses require more data than is routinely collected in therapy trials. A formula for sample-size and power for trials with a multiple-membership design will also be presented.

A trial of cognitive behavioural therapy for depression in adolescents that motivated this work will be used as an illustrative example.

Key words: Multiple Membership Models, Partial Nesting, Non-pharmacological Treatment Trials, Sample Size

---

*Health Sciences Research Group, School of Community Based Medicine , University of Manchester, UK*

# The clustered trend vector model


## Rooij, Mark de



We propose a new methodology to find clusters of participants with different trends for longitudinal categorical data. Standard random effects methodology for such data has serious computational and interpretational problems that are solved using the new procedure. For the estimation of the new method three variants of the Classiffcation EM Algorithm are described (one of them is deterministic, two are stochastic). An application of preferential choice of softdrink categories is presented showing all the merits of the clustered trend vector model.


.


Key words: Biplots; Categorical data; Longitudinal data; Random effects; Multidimensional Scaling.


---


*Mark de Rooij,Psychological Institute, Methodology and Statistics Unit, Leiden University. PO Box 9555, 2300 RB Leiden, The Netherlands. E-mail: rooijm@fsw.leidenuniv.nl*

# Factorial invariance in multilevel confirmatory factor analysis

**Ryu, Ehri**

With the development of multilevel structural equation modeling, multilevel confirmatory factor analysis is increasingly used in behavioral and social sciences. Factorial invariance can be tested at different levels of invariance: invariance in factor loadings, invariance in intercepts, and invariance in unique variances. Testing factorial invariance in two-level factor models requires considering invariance (at desirable level of invariance) at level 1 and at level 2. However not all levels of invariance are testable both at level 1 and level 2. In two-level factor models, the mean structure is typically captured at level 2 and the level-1 model includes no mean structure. When the group membership is at level 2, the level-1 mean structure is set to zero in both groups, and therefore it is not possible to test invariance in intercepts at level 1. When the group membership is at level 1, the group membership is crossed with the level-2 cluster. Two different ways of partitioning level-1 variables are considered: (a) level-2 cluster mean (combined across groups) and deviation from the cluster mean and (b) level-2 cluster by group mean (i.e., group-specific cluster mean) and deviation from the group-specific cluster mean. Implications of each partitioning and testable levels of invariance are discussed. The analyses will be demonstrated using artificial data.

Key words: factorial invariance, multilevel confirmatory factor analysis

---

*Ehri Ryu*
*Assistant Professor of Psychology*
*Department of Psychology*
*Boston College*
*Chestnut Hill, MA 02467*
*ehri.ryu.1@bc.edu*

# Modeling longitudinal MTMM data with interchangeable and structurally different methods: An application to developmental psychology.

**Schultze, Martin[1]; Koch, Tobias[1]; Geiser, Christian[2]; Eid, Michael[1]**

When assessing adolescents' social behavior longitudinally it is often useful to use several sources of information. Although self reports are often used, several studies have shown that especially peer reports contribute vital information. Because peer groups can be thought of as a population of individuals with access to similar information concerning an adolescent's behavior, they can be conceived as a population of *exchangeable methods*. Self reports, on the other hand, do not stem from this population of methods and are considered a *structurally different method* relative to peer reports. In this talk, an application of a newly developed multilevel CFA-MTMM model for longitudinal data collected from a combination of exchangeable and structurally different methods is presented (Koch et al., 2011). In the present study self reports and a varying number of peer reports of relational aggression and empathy were assessed in $N$ = 128 students at two measurement occasions. The statistical model presented here allows for the computation of variance components to quantify the degree of reliability, method specificity, and convergent validity between different methods. Results show that peers do not estimate adolescents' behavior in a homogeneous fashion (consistency ranged from .119 to .262) and their average rating shows low convergent validity with self reports (consistency ranged from .028 to .107). Additionally, self reports were less reliable (composite reliabilities ranged from .740 to .799) than peer reports (composite reliabilities ranged from .904 to .909). Both constructs were highly stable over a four month time period.

Key words: MTMM, structural equation modeling, multilevel modeling, exchangeable methods, structurally different methods, relational aggression, empathy, peer reports, self reports

---

[1] Freie Universität Berlin; [2] Arizona State University

Contact:
Martin Schultze
Freie Universität Berlin
FB Erziehungswissenschaft und Psychologie
AB Methoden und Evaluation
Habelschwerdter Allee 45
14195 Berlin
Germany
martin.schultze@fu-berlin.de

# Multilevel model with offset: A two stage approach for modeling respiratory prescribing, in Primary Health Care.

**Sofianopoulou, Eleni [⌂] *; Rushton Stephen [⌂] ; Pless-Mulloli Tanja ***

Asthma and Chronic Obstructive Pulmonary Disease (COPD) are two of the most common chronic respiratory diseases across the world. Exacerbation of respiratory symptoms can be triggered by air pollution as well as seasonal factors, such as pollen and respiratory infections. We aimed to develop a multilevel statistical model for assessing primarily the relationship between respiratory prescribing and air quality, at a primary health care setting. Lag periods between air pollution and respiratory prescribing were also examined.

We analysed respiratory medication (salbutamol) prescribed monthly by 64 primary care practices in Northeast of England, in 2002-2006. Firstly, we modelled the expectation of the area-wide monthly average respiratory prescribing rate, in relation to monthly average temperature, using a dynamic harmonic regression model. We then built the multilevel model using air quality, deprivation and demographic covariates per primary health practice, and including the monthly average respiratory prescribing rate, as offset.

The final model showed that an increase of 10μg/$m^3$ in ambient Particulate Matter ($PM_{10}$) concentrations was associated with an increase of 1% (95% CI, 0.1- 2%) in monthly salbutamol prescribing. The findings demonstrated that respiratory prescribing in primary care can be used as indicator of air pollution effect on asthma and COPD, increasing the scope of its use for health surveillance. Overall, the use of a multilevel model with offset provided a realistic representation of the relationship we assessed, as allowed to both handle the grouped nature of primary care data and account for the seasonal variation of respiratory health outcomes.

Key words: multilevel model, mixed-effects model, respiratory prescribing, air quality, Primary Health Care, harmonic regression

---

* *Institute of Health and Society, Newcastle University, UK*

[⌂] *Institute for Research on Environment and Sustainability, Newcastle University, UK*

# Detecting a missing level in a multilevel model and assessing its impact

**Spencer, Neil H.**

At the 7th International Amsterdam Multilevel Conference, some initial findings were presented by the author concerning the detection of misspecification in the multilevel model using the Forward Search (Atkinson & Riani, 2007). This earlier work relied on graphical representations to detect misspecification.

Work has progressed such that reliance on graphical methods is no longer required. The Forward Search is combined with simulations so that inferential methods can be used and misspecification in the form of a missing level in the hierarchy can be detected. To accomplish this, the level one residuals from a fitted multilevel model are examined. Summary statistics from the Forward Search are compared with what would be expected from a correctly specified model.

If a missing level is detected, the impact of this needs to be assessed. It is known that omitting a level may lead to a false impression being given of the importance of regressors. Additionally, the variation in the model will, necessarily, have been misallocated, potentially giving a false impression of the relative impact of different levels. To address this, different definitions of the missing level can be created using cluster analyses of the level one residuals. The multilevel model can then be refitted using these different definitions and the impact on the interpretation of the model assessed.

Demonstrations of the methods described above will be given using simulated and real data.

Atkinson, A.C. & Riani, M. (2007) “Exploratory tools for clustering multivariate data”, *Computational Statistics and Data Analysis*, 52, 272-285.

Key words: Misspecification; Forward Search; Cluster Analysis

---

*Email: N.H.Spencer@herts.ac.uk*
*Address: Business School, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, U.K.*

# A multilevel simultaneous equations model for within-cluster dynamic effects, with an application to reciprocal relationships between maternal depression and child behaviour

**Steele, Fiona**

We propose a general multilevel simultaneous equations model for studying dynamic interdependencies between individuals in a cluster, for example members of a family or employees in an organisation. The model includes lagged and cross-lagged terms to allow for effects of one individual's outcome at time t-1 on the change in another individual's outcome between times t-1 and t, while allowing for the possibility that the outcomes of members of the same cluster may be influenced by a common set of unmeasured time-invariant cluster characteristics. We show how the model can be framed as a type of multilevel multivariate model and estimated using multilevel modelling software.

Our approach is illustrated in an analysis of maternal depression and child delinquency using longitudinal data from the Avon Brothers and Sisters Study. We differentiate between reciprocal influences between a mother and her children, as well as reciprocal influences of siblings on one another, allowing effects to depend on characteristics of parent, child and sibling pair. The multilevel model allows for residual variation in mother and child outcomes at the occasion, individual and family levels, and residual correlation between mother and child outcomes due to shared unmeasured environmental and genetic factors. The model can also accommodate mixed family sizes.

# The use of multilevel modelling to minimise likely hidden bias in the design of matched control studies

**Steventon, Adam [1]* Richard Grieve [2]**

**Background**

Policymakers often need to understand the impact of a particular intervention, but a randomised control trial is not always feasible. Researchers often choose a matched control design where controls are chosen retrospectively using propensity scoring. A recurrent concern is that, although groups are similar in terms of what can be measured, they may nevertheless differ according to unobserved factors ("hidden bias").

We compared methods for selecting controls that take advantage of how interventions are often piloted in certain geographic areas. Selecting controls from outside of the area using a multilevel matching technique might result in estimates that are more robust to hidden bias.

**Methods**

We used simulations calibrated to a case study of hospital admissions to compare three methods: selecting controls from within the area that offered the intervention, from another arbitrarily-selected area, and from an area matched on area-level characteristics.

**Results**

Selecting controls from an arbitrarily-selected area resulted in large biases in the estimated treatment effect, for example if the unobserved individual-level covariate was distributed differently between areas.

Using a matched control area was found to improve balance on both observed and unobserved individual level variables, compared to selecting controls from within the area that offered the intervention. It also resulted in less biased estimate of the treatment effect. However, this assumed that variables related to the unobserved individual-level covariate are available to match on at the aggregate level.

**Conclusion**

When sufficient data are available, we recommend multilevel modelling at the design stage to determine the control area.

Key words: Multilevel modelling, design, evaluation, observational studies, propensity scores

---

*[1] The Nuffield Trust, 59 New Cavendish Street, London, UK*
*[2] London School of Hygiene and Tropical Medicine, Keppel Street, London, UK*

# A simple sample size formula for analysis of covariance in cluster randomized trials

**[1]Teerenstra, Steven; [2]Eldridge Sandra; [3]Graff, Maud; [1]de Hoop Esther; [1] Borm, George F.**

For cluster randomized trials with a continuous outcome, the sample size is often calculated as if an analysis of the outcomes at the end of the treatment period (follow-up scores) would be performed.
However, often a baseline measurement of the outcome is available or feasible to obtain. An ancova analysis using both the baseline and follow-up score of the outcome will then have more power.
We calculate the efficiency of an ancova analysis using the baseline scores compared an analysis on follow-up scores only. The sample size for such an ancova analysis is a factor $r^2$ smaller, where $r$ is the correlation of the cluster means between baseline and follow-up. This correlation can be expressed in clinically interpretable parameters: the correlation between baseline and follow-up of subjects (subject autocorrelation) and that of clusters (cluster autocorrelation). Because of that, subject matter knowledge can be used to provide a (range of) plausible values for these correlations, when estimates from previous studies are lacking.
Depending on how large the subject and cluster autocorrelations are, analysis of covariance can substantially reduce the number of cluster needed.

Key words: ANCOVA, pretest-posttest, cluster randomization, power, sample size

---

*[1]Departments of Epidemiology, Biostatistics and Health Technology Assessment, and [3]Scientific Institute for Quality of Healthcare, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*

*[2]Centre for Health Sciences, Barts and the London School of Medicine and Dentistry, Queen Mary, University of London, UK*

# Exploring the Performance of Information Criteria in Multilevel Structural Equation Modelling

**Ünal Çoker, Elif, Deniz Howe, Eylem & Howe, J. Andrew**

One of the problems in multilevel structural equation modeling (ML-SEM) is that of model selection. The problem of model selection in ML-SEM has not been studied in a wide range. Akaike's well-known model selection criteria, AIC, has been applied in the context of ML-SEM, but the effectiveness of many other information criteria have not been studied in a convincing manner. In this study, as a primary step in understanding the more complex ML-SEM, the information criteria in multilevel confirmatory factor analysis (ML-CFA) is examined. For this purpose, the performance of several AIC-type and ICOMP-type criteria are explored. To study the empirical performance of the information criteria, a simulation study is set up and it is demonstrated that most of the information-theoretic criteria select the correct model with very high frequencies.

Key words: Multilevel structural equation modelling, Multilevel confirmatory factor analysis, Information Criteria, AIC-type criteria, ICOMP-type criteria.

---

***Ünal Çoker, Elif[a], Deniz Howe, Eylem[a], Howe, J. Andrew[b]***

*[a] Mimar Sinan Fine Arts University, Faculty of Science and Letters, Department of Statistics, Istanbul, Turkey.*

*[b]TransAtlantic Petroleum, Istanbul, Turkey.*

# A Bayesian Multilevel IRT Model for Testing Measurement Invariance in Cross-national and Longitudinal Surveys

**Verhagen, Josine**

In large cross-national (e.g. PISA, TIMSS, SHARE, ESS, ISSP) or longitudinal surveys, data are structured hierarchically as persons within countries or observations at different time points within persons. Hierarchy is not only applicable to the scores on the survey, but also to the characteristics of the items in the survey, which can differ between countries or time points. By imposing a multilevel structure with random item effects on the (IRT) item parameters, it is possible to model these differences. In this way, tests for (measurement) invariance of these item characteristics can be performed without the usual need for anchor items. The only restriction is that a common shift of all item parameters within a group is not allowed. In addition, country- or time-specific covariates can be used to explain differences in item characteristics. Combined with a multilevel structure on the latent variable this offers a comprehensive model to analyze large cross-national or longitudinal surveys. Group differences in latent scores and in item parameters as well as fixed or random covariate effects explaining these differences can be estimated simultaneously with Bayesian MCMC methods. Examples of both cross-national and longitudinal applications will be given.

Key words: Bayesian inference, multilevel modeling, measurement invariance, cross-national research, longitudinal research.

---

*Josine Verhagen*
*University of Twente*
*Faculty of Behavioural Sciences*
*Department of Measurement and Data Analysis*
*P.O. Box 217, 7500 AE Enschede, The Netherlands*
*A.J.Verhagen@gw.utwente.nl*

# Approaches to Multilevel Repeated Measures Data: Roles for Centering

**West, Stephen**

Researchers in areas like developmental, health, and personality psychology often collect repeated measures data (level 1) on individuals (level 2), for example, in the form of daily dairies, event-contingent reports, or annual assessments. Two different general multilevel approaches may be taken depending on the research question. In the first approach, time is taken as a key variable and the focus is on linear or nonlinear trajectories of growth or decline in the data and their relationships to time-invariant (e.g., gender) or time-varying predictor variables. With some data systematic (e.g., weekly, monthly) cycles may also be of interest. Centering the data at one (or more) time points is used to provide a convenient interpretation of the results. In the second approach, time is ignored and the interest is in the within-person and between-person relationships with predictor variables. Centering in this approach helps partition the between- and within-person variation so that they can be studied separately, often yielding interesting relationships. These ideas will be illustrated with different empirical examples from substance use and educational achievement. Some links between multilevel modeling and structural equation modeling approaches to these research questions will also be made.