S4 Small Sample Size Solutions

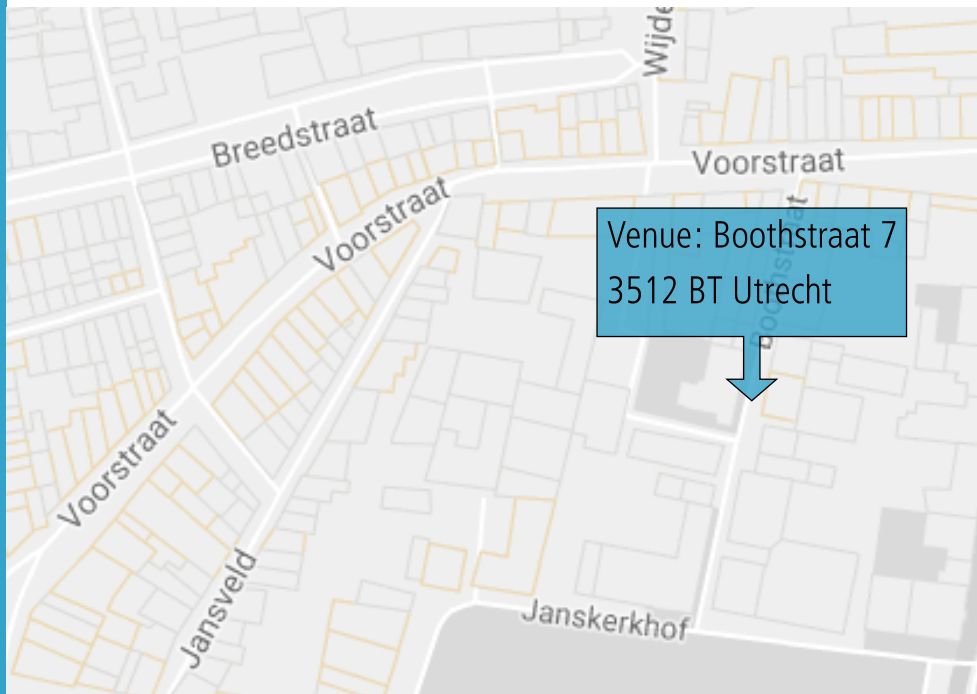
S4 Conference Programme:

Small Sample Size Solutions

Researchers often have difficulties collecting enough data to achieve adequate statistical power: when target groups are small, hard to access, or measuring the participants requires prohibitive costs. Such obstacles to collecting data usually lead to a limited data set. Researchers can overcome this through simplifying their hypotheses and statistical models. However, this strategy is undesirable since the intended research question cannot be answered in this way.

The aim of the S4 Conference is to bring together international researchers working to provide Solutions for Small Sample Size issues and to share information, learn about new developments and discuss solutions for typical small sample size problems

- **Date:** 5-8 March 2018
- **Venue:** Boothstraat 7, 3512BT, Utrecht
  Nearest bus stop: Janskerkhof
  (line 7,8,28,50,51,52,53,55,74,77)
- **Website:** www.uu.nl/s4
- **Programme (page 3-6):**
  - ⇒ **Monday:** Pre-conference workshops (page 7).
  - ⇒ **Tuesday:** Talks (page 11), poster presentations round 1(page 10), keynote speeches (page 9), dinner.
  - ⇒ **Wednesday:** Talks (page 11), poster presentations round 2 (page 10), keynote speeches, (page 9), young researcher award.
  - ⇒ **Thursday:** Post-conference workshops (page 8).

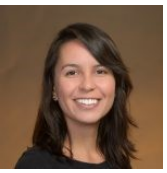


Utrecht University

S4 Small Sample Size Solutions

Organizing Committee

Rens van de Schoot
Associate Professor
Methodology & Statistics (Utrecht University)



Milica Miočević
Assistant Professor
Methodology & Statistics (Utrecht University)

Sarah Depaoli
Associate Professor
Quantitative Psychology (University of California)

Dan McNeish
Assistant Professor
Quantitative Psychology (Arizona State University)

Sanne Smid
PhD Candidate
Methodology & Statistics (Utrecht University)

Marianne Geelhoed
Organisation Committee
Methodology & Statistics (Utrecht University)

Laurent Smeets
Student Assistant
Methodology & Statistics (Utrecht University)

Utrecht University

Keynote Speakers

Dr. Joop Hox
Utrecht University
Small Data At Any Level? Problems and Solutions

Dr. Todd D. Little
Texas Tech University
On the Merits of Parceling

Dr. Marija Maric
University of Amsterdam
When less is more: Single-case research in youth clinical practice

Dr. Patrick Onghena
KU Leuven
One by one: The design and analysis of replicated randomized single-case experiments

Abstracts of the keynotes can be found on page 9

S4 Small Sample Size Solutions

Day by Day Programme

Day 1 – Monday March 5

9:00	**Registration pre-conference workshop 1**	
9:30	**Rens van de Schoot** ***pre-conference workshop – part 1***	*Gentle Introduction to Bayesian Analysis with Small Samples*
11:00	***Coffee and Tea Break***	
11:20	**Rens van de Schoot** ***pre-conference workshop - part 2***	*Gentle Introduction to Bayesian Analysis with Small Samples*
12:30	***Lunch***	
13:00	**Registration pre-conference workshop 2**	
13:30	**Milica Miočević** ***pre-conference workshop – part 1***	*Mediation Analysis with Small Samples*
15:00	***Coffee and Tea Break***	
15:20	**Milica Miočević** ***pre-conference workshop – part 2***	*Mediation Analysis with Small Samples*
16:30	**Conclusions day 1 and pre-conference workshops**	

Utrecht University

S4 Small Sample Size Solutions

Day 2 – Tuesday March 6

8:30	**Registration**	
9:00	**Opening by Rens van de Schoot**	
9:10	**Marija Maric**	*When less is more: Single-case research in youth clinical practice*
10:00	**Laslo Vincze**	*Problems with power in a mixed ANOVA*
10:20	**Alan Johnson**	*Ring out the old, Ring in the new: Field Research Designs for New Venture Teams using S4*
10:40	**Lucy Busija**	*When less is not more: Trials and tribulations of achieving the required sample size in a 'real-world' cluster RCT*
11:00	***Coffee and Tea Break***	
11:30	**Xynthia Kavelaars**	*Going multivariate in clinical trial studies: Increasing efficiency using Bayesian adaptive methods for information sharing*
11:50	**Fayette Klaassen**	*All for one or some for all? Bayesian evaluation of multiple N=1 hypotheses*
12:10	**Kimberley Lek**	*Extreme survival guide: what to do with a single person and few observations?*
12:30	***Lunch and poster round 1***	
13:30	**Patrick Onghena**	*One by one: The design and analysis of replicated randomized single-case experiments*
14:20	**Ana Slavec**	*Determining the number of replicates in experimental studies with wood samples: how low can we go?*
14:40	**Martina McMenamin**	*Improving the efficiency of rare disease trials using composite endpoints*
15:00	***Coffee and Tea Break***	
15:30	**Tina Nane**	*In and out of sample validation for structured expert judgment – a small sample size analysis*
15:50	**Leonard Vanbrabant**	*Sample-size reduction by order constraints*
16:10	**Herbert Hoijtink**	*Small Data is Becoming a Bigger Challenge*
16:50	**Conclusions day 2 by Milica Miočević**	
17:00	***Drinks (included)***	
17:30	***Dinner (registration is required)***	

Utrecht University

S4 Small Sample Size Solutions

Day 3 – Wednesday March 7

9:00	**Doors open**	
9:30	**Todd Little**	*On the Merits of Parceling*
10:20	**Jan de Neve**	*Regression models for rank tests when samples are small*
10:40	**Sanne Smid**	*Bayesian SEM with Informative Priors: Precautions and Guidelines*
11:00	***Coffee and Tea Break***	
11:30	**Marielle Zondervan**	*Searching prior information to solve small sample size issues in SEM*
11:50	**Marthe Egberts & Duco Veen**	*Increasing power of statistical analyses through collaboration*
12:10	**Duco Veen & Maartje de Klerk**	*Improving the Assessment of Individual Phoneme Discrimination Performance*
12:30	***Lunch and poster round 2***	
13:30	**Diana Zavala-Rojas**	*Bayesian Estimation of the True Score Multitrait–Multimethod Model with a Split -Ballot Design*
13:50	**Sara van Erp**	*Shrinkage priors for Bayesian penalized regression: An overview and tutorial using Stan*
14:10	**Caspar van Lissa**	*MetaForest: Exploring heterogeneity in meta-analysis using insights from machine learning.*
14:30	***Coffee and Tea Break***	
15:00	**Yves Rosseel**	*Small Sample Solutions for SEM*
15:50	**Joop Hox**	*Small Data At Any Level? Problems and Solutions*
16:40	**Young Researcher Award ceremony**	
16:45	***End day 3***	

Utrecht University

S4 Small Sample Size Solutions

Day 4 – Thursday March 8

9:00	**Registration post-conference workshop 1**	
9:30	**Joop Hox** ***post-conference workshop part 1***	*Multilevel Modeling with Small Samples*
11:00	***Coffee and Tea Break***	
11:20	**Joop Hox** ***post-conference workshop part 2***	*Multilevel Modeling with Small Samples*
12:30	***Lunch***	
13:30	**Registration post-conference workshop 2**	
13:30	**Rens van de Schoot** ***post-conference workshop part 1***	*Latent Growth Curve Modeling with Small Samples*
15:00	***Coffee and Tea Break***	
15:20	**Rens van de Schoot** ***post-conference workshop part 2***	*Latent Growth Curve Modeling with Small Samples*
16:30	**Conclusions day 4 and post-conference workshops**	

Utrecht University

Pre-conference Workshops

5 March

09.30-12.30

Gentle Introduction to Bayesian Analysis with Small Samples

Instructor: Rens van de Schoot

Bayesian methods are becoming increasingly popular as a solution for small sample problems in social sciences. Despite their popularity, these methods have yet to become a standard part of the statistics curricula in graduate programs.

This workshop is designed for applied researchers who are new to Bayesian methods and would like to learn the theory behind Bayesian statistics, the differences between Bayesian and frequentist statistics, and how to apply Bayesian methods to answer their research questions. The focus of the workshop will be on how Bayesian methods can be used in small sample research.

12.30-13.30

Break and Lunch

13.30-16.30

Mediation Analysis with Small Samples

Instructor: Milica Miočević

Mediation analysis is used to evaluate the mechanism through which the independent variable(s) affect the dependent variable(s). Bayesian methods with accurate prior distributions were found to increase power to detect the mediated effect in small samples.

The workshop starts with a tutorial on how to use Bayesian methods in linear models, and proceeds to cover two ways to do Bayesian mediation analysis. This workshop is designed for researchers who are familiar with the theory behind Bayesian statistics and linear regression analysis. Participants new to Bayesian methods are encouraged to first attend the workshop "Gentle Introduction to Bayesian Analysis with Small Samples".

Post-conference Workshops

8 March

09.30-12.30

Multilevel Modeling with Small Samples

Instructor: Joop Hox

Multilevel models (MLMs) are used to analyze data that have a nested (hierarchical) structure, e.g., students are nested within classrooms. In small samples MLMs encounter convergence issues and yield parameter estimates with poor statistical properties.

This workshop focuses on available methods for avoiding such issues. The target audience for this workshop are researchers in social sciences who work with nested data and small samples. Participants new to Bayesian methods are encouraged to first attend the workshop "Gentle Introduction to Bayesian Analysis with Small Samples".

12.30-13.30

Break and Lunch

13.30-16.30

Latent Growth (Mixture) Modeling with Small Samples

Instructor: Rens van de Schoot

Latent growth curve models (LGMs) and Latent Growth Mixture Models (LGMMs) are used to model the changes in a construct over time. In small samples such models often encounter convergence issues and yield parameter estimates with poor statistical properties.

Bayesian methods with informative prior distributions offer a solution to these issues. The target audience for this workshop are social sciences researchers who work with longitudinal data and small samples. Participants new to Bayesian methods are encouraged to first attend the workshop "Gentle Introduction to Bayesian Analysis with Small Samples".

Keynote Speeches

Dr. Joop Hox

Utrecht University

Small Data At Any Level? Problems and Solutions

The estimation methods commonly employed in multilevel analysis assume large sample sizes. Precisely what is a large enough sample size is unknown, but simulation studies suggest that having fewer than twenty groups definitely is a small sample of groups. Nevertheless, researchers encounter data where they need to analyze data from small samples.

In multilevel analysis, small samples can also be the result of having small groups. This is the case, for example, in dyadic research, where the unit of analysis is a couple. In such research, if there are some missing data, the average groups size can actually become lower that two. Also, in longitudinal research, it is common to have only a small number of measurement occasions, having only three measurement occasions is not uncommon.

This presentation reports the available evidence on sufficient sample sizes, and discusses analysis strategies that can mitigate the problems that occur with small sample sizes, such as estimation accuracy and statistical power.

Dr. Todd D. Little

Texas Tech University

On the Merits of Parceling

Parceling is a data pre-processing strategy by which two or more items are averaged to create a new aggregate indicator to use in both exploratory and confirmatory factor models (aka. Latent variable modeling, structural equation modeling). First introduced by Cattell over half a century ago, the practice of parceling has been a hotly debated practice and even earned the moniker "the items versus parcels controversy." In this lecture, I will outline the arguments both pro and con regarding the items versus parcels controversy. I will conclude with why the items versus parcels needn't be one and provide compelling reasons for why parcels are highly preferred.

Dr. Marija Maric

University of Amsterdam

When less is more: Single-case research in youth clinical practice

Single-case experimental designs (SCEDs) are increasingly recognized as a valuable alternative for Randomized Controlled Trials (RCTs) to test intervention effects in youth populations. Given the heterogeneous nature of youth and family problems, SCEDs may be the most optimal way to investigate intervention outcomes, either because the condition is rare (e.g., certain comorbidity) or because analyses on a group level would imply loss of information (i.e., finding no intervention effect while effect is present in a certain subgroup). The current presentation will provide an overview of a single-case method as a way to investigate effectiveness of youth interventions, along with the challenges related to assessment and data-analyses, accompanied by solutions and illustrations from single-case research with youth and families suffering from anxiety disorders, negative self-esteem, comorbid ADHD and anxiety disorders, and child abuse.

Dr. Patrick Onghena

KU Leuven

One by one: The design and analysis of replicated randomized single-case experiments

Single-case experiments are "true" experiments on single cases. In a single-case experiment, the experimental manipulation is introduced within-case by using several experimental phases or fast alternation of experimental conditions, and by using repeated measurements to assess the outcome. Randomization and replication can be included to strengthen the internal and external validity, respectively. Data analysis proceeds by paying attention to the specific characteristics of the single-case setup and the specific design that is used. A combination of structured visual analysis, calculation of effect size measures, causal inference, and meta-analysis seems most promising.

Poster Presentations

Day 2: Tuesday March 6

⇒ **Angulo-Brunet, Ariadna:** How to handle ordered responses with floor and ceiling effects in SEM using small samples?

⇒ **Declercq, Lies:** How can methodologists make multilevel modeling more accessible for applied researchers who use single-case experimental designs?

⇒ **Gibertoni, Dino:** How to correctly perform log-rank tests in a study population of 15 observations and less than 10 events?

⇒ **Jamshidi, Laleh:** The methodological quality of single-case experimental studies meta-analyses

⇒ **Moeyaert, Mariola:** How to Improve Bayesian Estimation When Synthesizing Single-Case Experimental Design Studies' Random Effects' Variance Components.

⇒ **Ruiter, Naomi de:** Studying temporal dependence within and between variables for (mixed-method) data with small samples?

⇒ **Soukup, Petr:** Inference for samples from small populations – Frequentist or Bayesian solution?

⇒ **Vrolijk, Paula:** How to deal with high attrition in a small complicated data set?

Day 3: Wednesday March 7

⇒ **Beretvas, Tasha:** How can we obtain unbiased ICC estimates for small sample size datasets?

⇒ **Dong, Shuyang:** Can latent growth modeling (LGM) and growth mixture modeling (GMM) be used with a sample size around 100?

⇒ **Gmelin, Ole:** How to deal with small sample sizes at the group-level for Dyadic Data Analysis in Speed-Dating Contexts?

⇒ **Langeloo, Annegien:** How to analyze a cross-lagged multilevel model to compare two small groups?

⇒ **Radulescu, Silvia:** How to analyze a sample with few data points per participant in a familiarization paradigm in artificial grammar learning?

⇒ **Song, Yue:** How to analysis the development of sharing across three waves, also the relationship between parenting and sharing during these waves, using a sample size around 90?

⇒ **Thauvoye, Evalyne:** Is a Bayesian approach the solution for care transition research? Beyond descriptive analysis in a difficult to reach and drop-out prone population.

Talks

⇒ **Busija, Lucy:** When less is not more: Trials and tribulations of achieving the required sample size in a 'real-world' cluster RCT. (page 12)

⇒ **De Neve, Jan:** Regression models for rank tests when samples are small. (page 13)

⇒ **Egberts, Marthe & Duco Veen:** Increasing power of statistical analyses through collaboration. (page 14)

⇒ **Erp, Sara van:** Shrinkage priors for Bayesian penalized regression: An overview and tutorial using Stan. (page 15)

⇒ **Hoijtink, Herbert:** Small Data is Becoming a Bigger Challenge. (page 16)

⇒ **Johnson, Alan R.:** Ring out the old, Ring in the new: Field Research Designs for New Venture Teams using S4. (page 17)

⇒ **Kavelaars, Xynthia:** Going multivariate in clinical trial studies: Increasing efficiency using Bayesian adaptive methods for information sharing. (page 18)

⇒ **Klaassen, Fayette:** All for one or some for all? Bayesian evaluation of multiple N=1 hypotheses. (page 19)

⇒ **Klerk, Maartje de & Duco Veen:** Improving the Assessment of Individual Phoneme Discrimination Performance. (page 20)

⇒ **Lek, Kimberley:** Extreme survival guide: what to do with a single person and few observations? (page 21)

⇒ **Lissa, Caspar van:** MetaForest: Exploring heterogeneity in meta-analysis using insights from machine learning. (page 22)

⇒ **McMenamin, Martina:** Improving the efficiency of rare disease trials using composite endpoints. (page 23)

⇒ **Nane, Tina:** In and out of sample validation for structured expert judgment – a small sample size analysis. (page 24)

⇒ **Rosseel, Yves:** Small Sample Solutions for SEM. (page 25)

⇒ **Slavec, Ana:** Determining the number of replicates in experimental studies with wood samples: how low can we go? (page 26)

⇒ **Smid, Sanne:** Bayesian SEM with Informative Priors: Precautions and Guidelines. (page 27)

⇒ **Vanbrabant, Leonard:** Sample-size reduction by order constraints. (page 28)

⇒ **Vincze, Laszlo:** Problems with power in a mixed ANOVA. (page 29)

⇒ **Zavala-Rojas, Diana:** Bayesian Estimation of the True Score Multitrait–Multimethod Model with a Split-Ballot Design. (page 30)

⇒ **Zondervan-Zwijnenburg, Mariëlle:** Searching prior information to solve small sample size issues in SEM. (page 31)

Utrecht University

When less is not more: Trials and tribulations of achieving the required sample size in a 'real-world' cluster RCT

Busija, L.*, Byers, J. McCabe, M
Institute for Health and Ageing, Australian Catholic University, Australia
**** Presenting author***

Summary

Cluster randomised controlled trials (CRCT) present researchers and statisticians with a number of special challenges. As a rule, CRCT necessitate larger sample sizes than individually randomised controlled trials, due to the clustering effect, which contributes to extra variation in the outcome. This extra variation in the outcome needs to be counteracted by a larger sample size, to ensure adequate power of the study. Statistically, CRCT with a large number of small clusters provides the most efficient design in terms of sample size required for the study. However, in 'real world' settings, delivering an intervention to a large number of organisations is inefficient from the point of view of resource allocation and can considerably increase the overall cost of a study. This case study describes the challenges in recruitment and maintenance of a sample size in a three-arm CRCT of the Resident at the Centre of Care (RCC) Program. The program was designed to assist staff of residential aged care facilities (RACF) transition to the model of care that was centred on the needs and choices of residents. The trial evaluated efficacy of the program alone or with on-going clinical support, relative to care as usual. Randomisation was at the level of RACF. The primary outcome of the program was resident quality of life, measured at the level of individual residents. Secondary outcome was improved organisational climate, measured at the level of RACF. Data were proposed to be analysed with a multilevel linear regression, with individuals clustered within RACF and random intercept for RACF. Sample size calculations for the primary outcome were performed to achieve 80% power and 5% Type I error rate (2-tailed). Input parameters for the sample size calculation were based on the results of our previous studies in RACF and assumed a moderately small effect of the intervention (Cohen's $d \geq 0.35$), intra-class correlation of 0.02, an average cluster size of 20 residents per RACF, and 25% attrition at 6 months. The resultant estimate for the total sample size at baseline was 744 residents and 39 RACF (13 for each arm of the trial). The estimated budget for the study was $1.3 million AUD. To date, we were able to secure $155,000 AUD, which allowed us to recruit and randomise 9 RACFs. With the expectation of 20 residents per RACFs, our smallest detectable effect size increased to $d=0.57$, hence reducing our chances of detecting a meaningful effect of the intervention. Additionally, the small number of RACFs recruited into the study means that the results of 'traditional' multilevel model for the secondary outcome (organisational climate at the level of RACF) are likely to have very low precision and power. Workable alternatives to the traditional large-sample multilevel models are urgently needed to increase feasibility of CRCT designs and optimise their implementation in 'real-world' settings.

Regression models for rank tests when samples are small

Jan De Neve[1]*, Olivier Thas[2,3] Gustavo Amorim[2], Karel Vermeulen[2] en Stijn Vansteelandt[4,5]

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

[2] *Department of Mathematical Modelling Statistics and Bioinformatics, Ghent University*

[3] *National Institute for Applied Statistics Research Australia (NIASRA),School of Mathematics and Applied Statistics, University of Wollongong*

[4] *Department of Applied Mathematics, Computer Science and Statistics, Ghent University*

[5] *Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine*

**** Presenting author***

Summary

We demonstrate how many classical rank tests, such as the Wilcoxon-Mann-Whitney, Kruskal-Wallis and Friedman test, can be embedded in a statistical modelling framework and how the method can be used to construct new rank tests. In addition to hypothesis testing, the method allows for estimating effect sizes with an informative interpretation, resulting in a better understanding of the data. Our method results from two particular parametrizations of probabilistic index models (Thas et al., 2012). The popularity of rank tests for small sample inference makes probabilistic index models also natural candidates for small sample studies. However inference for such models relies on asymptotic theory that can deliver poor approximations of the sampling distribution if the sample size is rather small. We therefore explore a bias-reduced version of the bootstrap and adjusted jackknife empirical likelihood and show that their application leads to drastic improvements in small sample inference for probabilistic index models. These results justify the use of such models for reliable and informative statistical inference in small sample studies.

Increasing power of statistical analyses through collaboration

Egberts, M.[1,2]*(PhD-student), Veen, D.[3]* (PhD-student), van de Schoot, R.[3] (supervisor), van Loey, N.[1,2] (supervisor)

[1] *Association of Dutch Burn Centres, Beverwijk, the Netherlands*

[2] *Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands*

[3] *Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Utrecht, The Netherlands*

Both Marthe and Duco will act as presenters during this presentation

Summary

Complex statistical models generally require large sample sizes. In practice, these numbers cannot always be easily obtained. In research on the psychological impact of pediatric burns on the family, problems with small sample sizes may arise. In order to obtain a sufficient sample size, prolonged multicenter studies are needed. Using informative priors to increase the power of the statistical analyses can be a solution to sample size problems. In the current project, we use informative priors to estimate a growth curve model with a distal outcome, using prospective data from mothers of young children with burns. The ultimate goal of the project is to compare results obtained with default priors, informative priors obtained from the literature, and priors resulting from expert knowledge. The presentation will be given by an applied researcher and a statistician, addressing issues faced when analyzing small datasets and potential solutions to these problems. Hopefully, this will provide insight in the way in which researchers of these two disciplines can collaborate and support each other.

Shrinkage priors for Bayesian penalized regression: An overview and tutorial using Stan.

Van Erp, S.[1]*, Oberski, D.L.[2], Mulder, J.[3]

[1] *Tilburg University, The Netherlands; PhD-student under supervision of Dr. Ir. J. Mulder, Dr. D.L. Oberski, and Prof. Dr. J.K. Vermunt*

[2] *Utrecht University, The Netherlands*

[3] *Tilburg University, The Netherland*

**** Presenting author***

Summary

This presentation focuses on the problem of a small sample size, relative to the number of predictors in a regression model. If the sample size is smaller than the number of predictors, the model is not identified and cannot be estimated using traditional regression approaches, such as ordinary least squares. Penalized regression methods such as the lasso (least absolute shrinkage and selection operator; Tibshirani, 1996) and the ridge (Hoerl & Kennard, 1970) are well-known solutions to this identification problem. The central idea of penalized regression approaches is to add an ad hoc penalty term to the minimization problem that will shrink small coefficients towards zero. However, it can be difficult to obtain valid standard errors and to determine the penalty parameter, which is a central parameter in these methods. These problems can be solved using a Bayesian approach in a straightforward manner. In a Bayesian analysis, a prior distribution is specified for all parameters, which, combined with the likelihood of the data, results in a posterior distribution. It is well known that specific shrinkage priors imply penalties equivalent to classical penalization methods, such as the ridge and lasso. Moreover, certain Bayesian penalization methods have been shown to perform similarly to or better than classical penalization methods (Hans, 2009; Kyung et al., 2010; Li & Lin, 2010), and additionally Bayesian methods straightforwardly result in credibility intervals with a clear Bayesian interpretation. Due to these advantages, Bayesian penalization is becoming increasingly popular and many different prior distributions have been proposed that have desirable properties in terms of prediction and variable selection. However, the extensive technical Bayesian literature and subtle differences between the priors can make it difficult for researchers to navigate the options and make sensible choices for the problem at hand. The goal of this talk is to aid researchers in this endeavor by presenting an overview and comparison of the different prior options and, importantly, providing insight in the characteristics and behaviors of the priors. Two methods will be presented to determine the penalty parameter: a full Bayesian method and an empirical Bayes method. All priors have been implemented in the freely available software package Stan (Stan development team, 2017), and all code will be made available online so that researchers can easily use the different priors.

Small Data is Becoming a Bigger Challenge

Herbert Hoijtink*
Methods and Statistics - University Utrecht
*** Presenting author**

Summary

This talk is based on Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.--J., … H., Hoijtink, …, & Johnson, V. E. (2017). Redefine Statistical Significance. Nature Human Behavior. DOI: 10.1038/s41562-017-0189-z.

In this paper the authors propose to change the default P-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.Three reasons for this are: 1) it reduces the probability of publication bias; 2) it reduces the options to use sloppy science to obtain significant test results; and 3) it corresponds to a Bayes factor of at least 10, that is, 10 times more (or less) support for the null versus the alternative hypothesis.

Of course using a smaller alpha-level has consequences for power. For example, to detect a medium effect size with a power of .80 in an independent samples t-test, 64 persons per group are sufficient if alpha equals .05. However, with alpha equal to .005, already 107 person per group are needed. Stated otherwise, small data is becoming a bigger challenge because traditional approaches to statistical inference will be hugely underpowered.

It is often thought that the use of informative hypotheses leads to a substantial increase in power. However comparing H0: both means are equal, versus. Ha: the first is larger than the second, still needs 50 persons per group (for alpha equals .05) and 94 persons per group (for alpha equals .005 and thus a corresponding Bayes factor of at least 10). Informative hypotheses not seem to be a convincing remedy for the power problem.

The talk will be concluded with two propositions that will be discussed with the audience:

1) We should all change the default alpha level from .05 to .005
2) Hypothesis evaluation by means of p-values or Bayes factors should not be used in case of small data

Ring out the old, Ring in the new: Field Research Designs for New Venture Teams using S4

Johnson, Alan R.[1, 2]* Delmar, Frédéric[3]
[1] *Nord University Business School, Bodø, Norway*
[2] *RATIO Research Institute, Stockholm, Sweden*
[3] *Lund University, Sweden*
*** Presenting author**

Summary

I thought to add an applied perspective on research design issues with clustered data and, hopefully, contribute to "shorten the chase" for other researchers based on our experiences with team level research questions and individual level observations (Johnson, van de Schoot, Delmar, & Crano, 2015).

My intention is to:

1) Start with our experiences squeezing results out of my dissertation data with 'traditional' repeated measures design using questionnaire instruments and a student sample of 60 new venture teams.
2) Then move on to Fred and my attempts to refining the external validity of that 'traditional' repeated measures design in Sweden still using questionnaire instruments with about 100 real new venture teams.
3) Draw attention to the resources needed and the expected response rates using field study designs in a less-controlled environment, i.e., not in a university but only in an entrepreneurship incubator/accelerator.
4) Consider some design alternatives:
   - big data designs considering issues of construct definition (Luciano, Mathieu, Park, & Tannenbaum, 2017) or
   - small data designs with 4 to 8 teams in a 2 by 2 matrix using intensive longitudinal designs (Kozlowski, 2015).
5) Division of labor between applied researchers and meteorologists:
   - applied researchers and their peer reviewers have a responsibility to be flexible and
   - meteorologists have a responsibility to provide alternatives that applied researchers can understand and evaluate appropriately.

References

- Johnson, A. R., van de Schoot, R., Delmar, F., & Crano, W. D. (2015). Social influence interpretation of interpersonal processes and team performance over time using Bayesian model selection. Journal of Management, 41(2), 574–606. https://doi.org/10.1177/0149206314539351
- Kozlowski, S. W. J. (2015). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. Organizational Psychology Review, 5(4), 270–299. https://doi.org/10.1177/2041386614533586
- Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2017). A Fitting Approach to Construct and Measurement Alignment: The Role of Big Data in Advancing Dynamic Theories. Organizational Research Methods, 1094428117728372. https://doi.org/10.1177/1094428117728372

Going multivariate in clinical trial studies: Increasing efficiency using Bayesian adaptive methods for information sharing

Xynthia Kavelaars[1]*, Maurits Kaptein, and Joris Mulder
Tilburg University
*** Presenting author**

Summary

Randomized controlled clinical trials (RCTs) are the gold standard to investigate the effectiveness of treatments for (mental) diseases. However, demonstrating the superiority of a treatment is often challenging, in part due to difficulties to enroll a sufficient number of patients. Small samples might lead to premature stopping of the trial, inconclusive results, and potentially the withholding of better treatments from patients in need.

Despite the apparent lack of information in small samples, we often have other sources of information available that could improve decision-making. Especially Bayesian adaptive trial designs lend themselves perfectly to take advantage of combining information from multiple complementary outcomes. Bayesian methods potentially allow for modeling complex dependency structures of multiple outcomes, learning the treatment effects using Bayesian updating, performing optional stopping when evidence is conclusive, and tailoring treatments to patients during the trial; all without compromising error rates.

However, (Bayesian) statistical methods that take advantage of information sharing between multiple outcomes in clinical studies are largely underdeveloped. In this presentation Bayesian (adaptive) methods are proposed for this purpose. Our study demonstrates how the Bayesian adaptive approach benefits from combining multiple outcomes allowing trials to stop earlier, possibly rendering a small sample sufficiently informative.

All for one or some for all? Bayesian evaluation of multiple N=1 hypotheses

Klaassen, F.[1]* (PhD student), Zedelius, C. M.[2], Veling, H.[3], Aarts, H.[4] and Hoijtink, H.[1,5] (supervisor)

[1] *Department of Methodology and Statistics, Utrecht University, The Netherlands*

[2] *Department of Psychology, University of Santa Barbara, USA*

[3] *Behavioural Science Institute, Radboud Universiteit Nijmegen, The Netherlands*

[4] *Department of Psychology, Utrecht University, The Netherlands*

[5] *Cito Institute for Educational Testing, Arnhem, The Netherlands*

*** Presenting author**

Summary

Analyses are mostly executed at the population level, whereas in many applications the interest is on the individual level instead of the population level. This research project considers multiple N=1 experiments, where participants perform multiple trials with a dichotomous outcome in various conditions. Expectations with respect to the performance of participants can be translated into so-called informative hypotheses. These hypotheses can be evaluated for each participant separately using Bayes factors. A Bayes factor expresses the relative evidence for two hypotheses based on the data of one individual. We propose to "average" these individual Bayes factors in the gP-BF, the average relative evidence. The gP-BF can be used to determine whether one hypothesis is preferred over another for all individuals under investigation. This measure provides insight into whether the relative preference of a hypothesis from a pre-defined set is homogeneous over individuals. Two additional measures are proposed to support the interpretation of the gP-BF: the Evidence Rate (ER), the proportion of individual Bayes factors that support the same hypothesis as the gP-BF, and the Stability Rate (SR), the proportion of individual Bayes factors that express a stronger support than the gP-BF. These three statistics can be used to determine the relative support in the data for the informative hypotheses entertained. Software is available that can be used to execute the approach proposed and to determine the sensitivity of the outcomes with respect to the number of participants and within condition replications.

Improving the Assessment of Individual Phoneme Discrimination Performance

De Klerk, M.K.A.[1]*, Veen, D.[2]*, de Bree, E.[3], & Wijnen, F.[1]

[1] Utrecht Institute of Linguistics OTS, Humanities Faculty, Utrecht University, Utrecht, The Netherlands

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

[3] Department of Developmental Disorders and Special Education, aculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands

Both Maartje and Duco will act as presenters during this presentation

Summary

Through research in infant development, substantial insight has been gained in all kinds of (non-linguistic) developmental processes and changes. For instance, we know that newborns start out life as language-general listeners and become language-specific listeners. Such findings, however, are based on group results that often include high variation in listening times and a limited number of test trials. This makes it difficult to interpret the data, especially when small sample groups are used. In the current study with 6 to 10-month-old Dutch infants a design is used that previously allowed the assessment of discrimination of different speech stimuli by infants. In contrast to the previous findings, in our study the individual analyses using a linear regression model with autoregressive (AR1) error structure did not yield the expected results of finding discrimination of speech stimuli. This whilst at a group level the effect was found. We show an alternative way of analyzing the data by means of a Hierarchical Bayesian analysis so that we can assess the group level effects and the individual effects simultaneously. By doing so we can include information from the group level in the individual assessments thereby substantially reducing the noise so that we can gain a better perspective on whether the individual infants can discriminate different speech stimuli.

Extreme survival guide: what to do with a *single* person and few observations?

Lek, K. M.[1]* (PhD student), Van De Schoot, R.[1] (supervisor)
[1] *Utrecht University, the Netherlands*
***Presenting author**

Summary

Sometimes, our interest is not in the common denominator of a group of people (i.e., 'nomothetic') but in a *specific* person (i.e., 'idiographic'). An example can be found in education and psychology, where we might be interested in estimating the ability of a particular student. Whereas our 'standard statistical toolbox' is suited for most nomothetic research, idiographic research asks for an update of this toolbox. *Especially* when information about this single person is scarce.

During this presentation, we explore the possibility of Bayesian statistics to estimate a student's (mathematical) ability based on only one or a few test results. We show how sources of uncertainty of the test (i.e., measurement error, uncertainty in the translation of raw test scores to percentile scores, statistical ties) can be incorporated in the Bayesian analysis.

Additionally, we discuss the possible merits of specifying subjective priors for specific persons based on information other than the test results. An interesting but controversial example are observations of the teacher of a single student. During the presentation, we illustrate how teacher insights about specific students can be elicited and how these elicited insights can be used as a prior in the Bayesian analysis of a specific student's abilities.

MetaForest: Exploring heterogeneity in meta-analysis using insights from machine learning.

Lissa, van C. J.[1]*

[1] *Erasmus University Rotterdam, NL*

*** Presenting author**

Summary

Meta-analysis often presents a small sample problem: The number of studies on any given topic is typically low, because conducting research is cost- and time-intensive. Human behavior, however, is notoriously complex (Earp & Trafimow, 2015), and consequently, subject to a host of potential moderators (Cesario, 2014). Additional moderators are introduced because similar research questions are examined in different labs, sampling from different populations, using idiosyncratic methods and instrumentation (Higgins et al., 2009). Even replication studies, by definition designed to be equivalent, typically display heterogeneity due to unforeseen moderators (Maxwell et al., 2015; Simmons et al., 2011). Finally, the paucity of theory regarding sources of heterogeneity at the between-studies level, makes it hard to whittle the list of potential moderators down to a manageable number (Thompson & Higgins, 2002). Meta-analysts are thus faced with what is known as the "curse of dimensionality": The problem that arises when the number of variables to be considered is large, relative to the number of cases in the data. Such cases do not fit comfortably into the classic meta-analysis paradigm, which, like any regression-based approach, requires many cases per parameter. This may partly explain why, despite the fact that software to conduct meta-analysis with multiple moderators is readily available (Viechtbauer, 2010), most published meta-analyses do not account for more than a few moderators, if any. In many cases, the sample size is simply too low to obtain the power required to reliably examine heterogeneity (Riley, Higgins, & Deeks, 2011).

Three approaches have been proposed to deal with between-studies heterogeneity (Higgins et al., 2009): First, if studies are assumed to be different, they should not be meta-analyzed. Secondly, if they are similar, a random-effects model can estimate the distribution of the true effect size. Thirdly, if known differences between studies introduce heterogeneity, these moderators can be accounted for using meta-regression. What is currently lacking is a "fourth approach", for cases where heterogeneity is suspected, but the causes are unknown. This calls for an exploratory technique which can perform variable selection — indentifying which moderators most strongly influence the observed effect size.

MetaForest aims to address this need. This technique applies random-effects weights from classic meta-analysis to random forests' bootstrapping procedure. Random forests are a powerful learning algorithm, flexible yet relatively robust to overfitting. Simulation studies show that, even in datasets as small as 20 cases, MetaForest has excellent performance in terms of three metrics: 1) Predictive performance, in terms of cross-validated R^2_{cv}; 2) power, as evidenced by the proportion of datasets in which the algorithm achieved a positive R^2_{cv}; and 3) the ability to distinguish relevant moderators from irrelevant moderators, using variable importance measures. My presentation will cover these simulations, and provide a short tutorial. Although MetaForest constitutes a fully-fledged paradigm for meta-analysis, it can also be integrated in existing workflows, as a final check to ensure that important moderators have not been overlooked. We hope that this approach will be of use to researchers, and that the convenient R package will facilitate its adoption. Please install.packages("metaforest")

Improving the efficiency of rare disease trials using composite endpoints

Martina McMenamin[1]* (PhD student), Anna Berglind[2], James Wason[1,3]

[1] *MRC Biostatistics Unit, University of Cambridge, Cambridge, UK*

[2] *Global Medicines Development, Biometrics & Information Sciences, AstraZeneca, Sweden*

[3] *Institute of Health and Society, Newcastle University, Newcastle, UK*

*** Presenting author**

Summary

Composite endpoints combining continuous and binary measures into a single endpoint are common in clinical development, particularly in autoimmune diseases and solid tumour oncology. It is well acknowledged in the literature that these endpoints can be useful for trials in rare diseases. By allowing more than one event to indicate effectiveness of a treatment, we ensure the complexity of many rare disease manifestations are captured. Furthermore, the higher event rates that result from combining single measures often translate to the requirement for smaller samples.

Standard practice when assessing the performance of a treatment in these settings is to dichotomise the information recorded on the continuous scale and combine this with the binary measures. This provides a single binary outcome based upon whether patients reach a predefined goal, i.e. are 'responders'. Dichotomising continuous variables is highly statistically inefficient. This is particularly problematic in disease areas with few patients, such as Lupus Nephritis, which struggle to recruit the required sample size for a clinical trial. An alternative, originally proposed by (Wason & Seaman 2013), is the augmented binary method. It employs joint modelling techniques which retain information on how close patients are to being 'responders'. It has been demonstrated to result in substantial efficiency gains when applied to phase II cancer trials and in OSKIRA-1, a phase III trial in rheumatoid arthritis. However, in these cases, the sample sizes considered were much larger than would be possible in many rare diseases.

We aim to determine whether these gains are also experienced in smaller samples. Previous work suggested there may be problems with the augmented binary method in small samples due to an increased number of parameters. We evaluate the behaviour of the augmented binary method in terms of type I error rate, power and coverage when we have few available patients ($n<100$), by resampling from the OSKIRA-1 trial. We identify finite sample corrections and implement these in the augmented binary method to improve its small sample properties. We compare this with the operating characteristics of the standard binary method and show that the augmented binary method with small sample corrections maintains nominal type I error rate (5%) whilst still offering much higher power. We make recommendations for future evaluations of treatments in rare diseases that utilise these endpoints.

In and out of sample validation for structured expert judgment – a small sample size analysis

Tina Nane[1*]
TU Delft
*** Presenting author**

Summary

The Classical Model or Cooke's method for structured expert judgment is a well-established approach in eliciting probability distributions. Its trademark is the presence of the calibration questions, which enable the quantification of experts' statistical accuracy and informativeness. The two scores yield, in turn, performance-based weights which enable a mathematical aggregation of experts' assessments to a so-called decision maker. The Classical Model has been used in numerous applications, which span from predicting possible malfunctions of chemical installations for the accident consequence management for nuclear power plants to attribution of global foodborne disease to specific foods.

Numerous studies have compared the Classical Model's performance-based weighting with equal-based or harmonic weighting. The performance-based weighting is usually shown to improve the information score of the decision maker, while preserving the calibration score. Moreover, the performance-based weighting has been extensively validated, both in sample and out of sample, using an updated TU Delft structured expert judgment database, from 2006 until 2015. The Classical Model can produce robust results for a reasonable small number of calibration questions and for a limited number of experts. I will present results from the updated TU Delft database, as well as from recent applications.

Small Sample Solutions for SEM

Yves Rosseel[1]*

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

*** Presenting author**

Summary

In this presentation, an overview will be given of old and recent solutions to handle small samples in the framework of structural equation modeling (SEM). A distinction is made between (1) solutions for estimation (and avoiding non-convergence issues), and (2) solutions for small-sample inference.

For estimation, we will advocate the use of a divide-and-conquer approach. The general (and old) idea is to break down the model into smaller pieces, estimate the parameters of each piece in turn, and finally combine these pieces again to get the final result. For example, in a SEM with many (measured) latent variables, one can estimate the parameters of each measurement model, one at a time. Once all the measurement parts have been estimated, we can hold the parameters fixed to these estimates, and only estimate the parameters of the structural part in a second step. Another approach is to first generate factor scores for each latent variable, and then (after a suitable transformation, known as the Croon correction) use these factor scores as if they are observed variables in a path analysis. Remarkably, both approaches are able to obtain consistent estimates. The end result is that we can estimate fairly large models, with a relatively small sample, and still get stable results in a frequentist framework.

For inference, we will discuss several methods to obtain unbiased standard errors, if we use a divide-and-conquer approach. As expected, the price to pay for using a multiple-step approach (instead of a single-step approach) is a modest loss of efficiency. But even if we use a classic SEM to estimate all the parameters in a single step, corrections are needed to get standard errors (and confidence intervals) that lead to better small-sample behavior. For the goodness-of-fit test statistic, we will discuss a Bartlett correction, including extensions to the non-normal case, and the incomplete data case.

Finally, if time permits, we will briefly discuss extensions of these small-sample solutions to the multilevel SEM setting.

Determining the number of replicates in experimental studies with wood samples: how low can we go?

Slavec, Ana[1]*, Burnard, Michael[1,2], Schwarzkopf, Matthew[1,2]

[1] *InnoRenew CoE, 6310-Izola/Isola, Slovenia*

[2] *University of Primorska, Andrej Marušič Institute, 6000-Koper, Slovenia*

*** Presenting author**

Summary

When designing experiments with wood products, it is important to provide a high enough number of replicates as wood specimens, even if taken from the same tree, may differ in their physical properties. On the other hand, conducting experiments with many wood specimens is time consuming, expensive, and researchers want to keep the cost and waste to the minimum. To determine the optimal number of replicates needed in experiments, an estimate of variance obtained from previous experiments is needed. However, in the field of wood science, these previous experiments with a comparable design are often rare, undocumented, or non-existent. Thus, the decision on the number of replicates is usually based on limited evidence. Sometimes researchers in wood science limit their replicate number to five or fewer.

To improve the design of further experiments in wood science, we analysed the data of two previous experimental studies in wood science that used 8 and 10 replicates, respectively. The first study was an investigation of the impact of different manufacturing decisions in the production of plywood panels with decorative veneers. As an indicator of quality, the development of cracks in the veneer was measured using the digital image correlation (DIC) technique. Four manufacturing factors were examined: veneer type (4 levels), core type (4 levels), adhesive type (3 levels), and lathe check orientation (2 levels), resulting in 96 combinations, each replicated 8 times.

The second study tested chemically modified lampante oil as a potential wood preservative. In this study, weight changes were recorded after leaching tests. Two factors were examined: wood species (2 levels) and modification treatment used (4 levels), resulting in 8 combinations, each replicated 10 times. For both studies the bootstrap resampling technique was used to generate subsamples with a lower number of replicates. Specifically, we observed the effect of reducing the size to 6 and 4 replicates in the first study and to 8, 6, and 4 replicates in the second study. We observed the increase in error in the reduced samples and the effects on the results of the two experiments. In addition, we compared it to the decrease of costs and calculated the optimal number of replicates for both. The results of the analysis will help researchers to optimize the design of further experiments that involve wood products.

Bayesian SEM with Informative Priors: Precautions and Guidelines

Smid, S.C.[1]*(PhD student), Depaoli, S.[2], McNeish, D.[3], Miočević, M.[1], Van de Schoot, R. (supervisor).[1,4]

[1] *Department of Methods and Statistics, Utrecht University, The Netherlands*

[2] *Department of Psychological Sciences, University of California, Merced, USA*

[3] *Psychology Department, Arizona State University, USA.*

[4] *Optentia Research Focus Area, North-West University, South Africa*

*** Presenting author**

Summary

Bayesian estimation is frequently suggested as a viable estimation method in small sample contexts. In a systematic literature review, we investigated the legitimacy of broadly applying Bayesian methods to address small sample sizes for structural equation models, instead of using frequentist methods (e.g. maximum likelihood estimation). Based on this review, we concluded that Bayesian estimation requires the inclusion of prior information for it to perform well with small samples. In fact, the use of only default (i.e., diffuse) priors can cause more bias than with frequentist methods, especially for the variance parameters in the model.

This conclusion raised a lot of new questions: e.g. is prior information required on all parameters? Likewise, how informative should the distribution(s) be? In a simulation study, we therefore investigated the performance of Bayesian estimation with varying combinations of informative and non-informative priors for a latent growth curve model with a distal (long-term) outcome under conditions with small samples. We varied the level of informativeness of the prior distributions, as well as the parameters on which the informative priors were placed. A selection of the informative prior distributions was specified in a way that can also be used in practice. As a result, applied researchers can easy incorporate our findings into their own work.

The goal of the simulation study is to find out which parameters require prior information, how informative these prior distributions should be to obtain accurate results, and what happens when prior distributions are specified that deviate from the true population values.

Sample-size reduction by order constraints

Vanbrabant, L.[1,2]* (PhD student, supervisor: Y. Rosseel), Kuiper, R.[2]

[1] Department of Data Analysis, Ghent University

[2] Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University

*** Presenting author**

Summary

Researchers often have substantive research questions that involve informative hypotheses. Consider, for example the hypothesis that cognitive behavioral therapy (CBT) in combination with drugs is more effective against depression than CBT only. This hypothesis is called informative because it includes a directional expectation about the ordering of the parameters. This prior knowledge originates from previous research (i.e. theory) or academicreasoning and can be translated into an order-constrained hypothesis by means of imposing order constraints (i.e. $\leq$, $\geq$, $=$) on the model parameters. Thus, in statistical symbols this informative hypothesis might be expressed as the following order-constrained hypothesis H1: $\mu_{\text{new drug}} \leq \mu_{\text{old drug}} \leq \mu_{\text{no drug}}$, where μ reflects the population mean for depression in each group.

In this presentation, we present a method to evaluate an informative hypothesis against its complement Hc, (Hc: not H1) using the generalized order-restricted information criterion (GORIC). Confirmatory approaches such as the GORIC have proven to be more `powerful' (higher probability of choosing the best hypothesis) than exploratory approaches such as the AIC. Consequently, smaller samples are needed to detect effects. In addition, we show how an informative hypothesis can be evaluated against its complement using the available software tools.

Problems with power in a mixed ANOVA

Laszlo Vincze[1]*

[1] *Swedish School of Social Science, University of Helsinki*

*** Presenting author**

Summary

The purpose of the present study is to establish the longitudinal change in levels of depression among a group of Finnish mothers with high levels of depression before giving birth to their first child. Also, a further purpose is to ascertain whether the change varies among married and cohabitating mothers. Questionnaire data was collected at three time points: three months before the birth of the baby, six months after the birth, and twelve months after the birth. Due to missing cases, there was a gradual decrease in sample size, $N_1 = 25$, $N_2 = 20$, and $N_3 = 15$.

Using standard statistical methods, the final sample in the analysis consists $N_3 = 15$ participants due to case-wise deletion; thus inferences can only be made about those who filled in the questionnaire on all three occasions. In general, married mothers (n = 7) reported higher levels of depression than cohabitating mothers (n=8). The mean difference between the two groups was increasing over time; it was 2.4 at the first time point, 3.93 at the second one and 4.83 at the third one.

Conducting two within-subject ANOVA separately among cohabitating and married mothers indicated significant change across time points among cohabitating mothers, $F = 19.46$, $p < .001$, $\eta_p^2 = .74$. However, the change over time was not significant among married mothers even if the effect size was considerably large, $F = 2.32$, $p = .14$, $\eta_p^2 = .28$.

A 3x2 mixed ANOVA demonstrated a non-significant interaction between time and group ($p = .64$) because of low power (.11); yet, the effect size for the interaction effect was meaningful $\eta_p^2 = .07$. The between subject effect of group was not significant either ($p = .12$).

Bayesian Estimation of the True Score Multitrait-Multimethod Model with a Split-Ballot Design

Diana Zavala-Rojas[1]*, Laura Castro-Schilo[2], Jonathan Lee Helm[3], Anna DeCastellarnau[3,4] and Zita Oravecz[5]

[1] *Universitat Pompeu Fabra, Spain*
[2] *SAS Institute, Inc., USA*
[3] *San Diego State University, USA*
[4] *Tilburg University, Netherlands. PhD-student, supervisor: Jeroen K. Vermunt*
[5] *The Pennsylvania State University, USA*
*** Presenting author**

Summary

This article (Helm et al. 2017) examines whether Bayesian estimation with minimally informative prior distributions can alleviate the estimation problems often encountered when fitting the true score multitrait–multimethod (MTMM) structural equation model with 2-group split-ballot data. This data design has been used in seven rounds of the European Social Survey (ESS). The split-ballot approach has been combined with the MTMM model to reduce the response burden of the respondents that would only need to answer 2 times to the same question, instead of 3. The 2-group split-ballot design, was implemented in the ESS to maintain, for substantive purposes, the total sample size for the first method, with a planned missingness design.

In particular, the 2-group true score MTMM structural equation model encounters an empirical under identification when (a) latent variable correlations are too homogenous, and (b) fitted to data from a 2-group split-ballot design. Monte Carlo simulations showed that problems were especially important when the sample size of the experiments was small (Revilla & Saris, 2013). Problems were not present with large datasets (20,000 cases), however, in practice it is very difficult to get huge sample sizes in survey research.

In this article we show, with Monte Carlo simulations and 3 empirical examples, that Bayesian estimation performs better than maximum likelihood estimation.

References:

- Helm, J. L., Castro-Schilo, L., Zavala-Rojas, D., DeCastellarnau, A., & Oravecz, Z. (2017). Bayesian Estimation of the True Score Multitrait–Multimethod Model With a Split-Ballot Design. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–15.
- Saris, W. E., Satorra, A., & Coenders, G. (2004). A New Approach to Evaluating the Quality of Measurement Instruments: The Split Ballot MTMM Design. *Sociological Methodology*, 34(1), 311–347.
- Revilla, M., & Saris, W. E. (2013). The Split-Ballot Multitrait-Multimethod Approach: Implementation and Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(1), 27–46.

Searching prior information to solve small sample size issues in SEM

Zondervan-Zwijnenburg, M.A.J.[1]* (PhD student), Peeters, M.[1], Depaoli, S.[2], Van de Schoot, R.[1,3] (supervisor)
[1] *Utrecht University, The Netherlands*
[2] *University of California, Merced, USA*
[3] *North-West University, South Africa*
*** Presenting author**

Summary

In this presentation, we will demonstrate guidelines on how to search and specify prior information for parameters in a structural equation model. We will demonstrate the guidelines by means of an empirical application about development of working memory in young heavy cannabis users (n = 16) and non-using peers (n = 252). To obtain prior information for the latent growth curve model of interest, meta-analyses, reviews, empirical papers and experts were involved. We will explain our systematic approach, comment on our experiences, and provide general recommendations to assist researchers that want to incorporate prior knowledge in a structural equation model.

Reference

Zondervan-Zwijnenburg, M. A. J., Peeters, M., Depaoli, S., & Van de Schoot, R. (in press). Where do priors come from? Applying guidelines to construct informative priors in small sample research. Research in Human Development. doi: 10.1080/15427609.2017.1370966