Abstract Title Page
Not included in page count.

Title: Multilevel synthesis of single-case experimental data: An empirical validation

Authors and Affiliations:
Van den Noortgate, Wim[1] (Wim.VandenNoortgate@kuleuven-kortrijk.be)

Moeyaert, Mariola[1] (Mariola.Moeyaert@kuleuven-kortrijk.be)

Ugille, Maaike[1] (Maaike.Ugille@kuleuven-kortrijk.be)

Beretvas, Tasha[2] (tasha.beretvas@mail.utexas.edu)

Ferron, John[3] (Ferron@usf.edu)

[1] Katholieke Universiteit Leuven, Belgium
[2] University of Texas at Austin
[3] University of South Florida

Acknowledgment:
This study is performed with a grant of the Institute of Education Sciences (IES Grant number R305D110024)

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Single-case or single-subject experimental designs (SSED) are used to evaluate the effect of one or more treatments on a single case. The basic interrupted time series design has a baseline phase consisting of a series of observations prior to treatment introduction, and a treatment phase consisting of a series of observations following treatment introduction.
Although SSED studies are growing in popularity, the results are in theory case-specific. To enhance generalizability, researchers can replicate the design across cases, either across studies, or within a primary study such as through the application of a multiple-baseline design (MBD), including interrupted time-series data from multiple participants where the timing of the intervention is staggered across the series. By synthesizing SSED studies' results, we can investigate the overall effect of an intervention, explore the generalizability of this effect, and look for factors that moderate the effect.

One systematic and statistical approach for combining single-case data within and across studies is multilevel modeling (Nugent, 1996; Shadish & Rindskopf, 2007; Van den Noortgate & Onghena, 2003a, 2003b, 2008). Multilevel models were developed for analyzing clustered data, such as data collected from students that can be grouped or are nested in schools. In SSED data from multiple cases, such as in a MBD study, we also have a nested structure. First, cases are measured repeatedly under different conditions. These measurements are grouped or 'nested' in subjects. If we have several SSED studies, subjects in turn are grouped in studies, meaning that three hierarchical levels can be distinguished.

Use of the multilevel modeling framework provides an appealing option because it can be used to provide estimates of individual treatment effects and how these effects change over time, estimates of the average treatment effect and how this effect changes over time, estimates of the variability in treatment effects, and estimates of the effects of moderators on the treatment effect and on the pattern of a treatment's effects over time. In addition, the models are flexible enough to handle (a) the nesting of observations and of outcomes within cases and the nesting of cases within studies, (b) a variety of forms for the growth trajectory within each phase of the design (e.g., linear, curvilinear), (c) alternative dependent error structures for the growth trajectories (e.g., first order autoregressive, toeplitz), (d) heterogeneous variances (within cases, across cases, or across studies), and (e) different types of outcomes (e.g., continuous, count), and (f) standardized or unstandardized raw data or effect size measures.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The purpose of the study is to investigate empirically the multilevel approach for combining SSED data. More specifically, we aim at assessing the value of the approach for numbers of observations, cases and studies that are common in SSED research, by looking at the bias and precision of the parameter estimates, and the validity of statistical inferences. Special

attention is given to the problem of standardization: if data are not measured on the same scale in each of the studies, they should be standardized before being combined. An option to make linearly equitable scales with a meaningful zero value comparable, is to divide the scores for each case by (an estimate of) the within phase standard deviation. The value of this approach has to be investigated yet.

With the obtained study results, we want to inform applied researchers about possibilities and limitations of the use of the basic multilevel model for combining SSED data. At the same time, the results will give indications about conditions for which the model or the estimation procedures should be further developed (e.g., by using bootstrap procedures instead of maximum likelihood procedures).

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

Although the multilevel modeling approach and its flexibility are appealing, there is much about SSED data and the functioning of multilevel modeling with this type of data and design that is not fully understood. For example, how well are parameters and standard errors estimated for the typical sample sizes encountered in SSED research? What happens to the accuracy of our inferences when the model is not correctly specified? How should we standardize our measures of effect? How many cases/studies are required to make valid inferences and to have adequate power?

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

The model that will be investigated in this study is a multilevel extension of the model of Center, Skiba, and Casey (1985-1986). More specifically, the observed scores for case *j* from study *k* are regressed on a time indicator, *T*, that is centered around the first observation of the intervention phase, a dummy variable for the treatment phase, and an interaction term of these variables:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} T_{ijk} + \beta_{2jk} D_{ijk} + \beta_{3jk} T_{ijk} D_{ijk} + e_{ijk}$$

The equation shows that the expected score in the baseline phase equals $\beta_{0jk} + \beta_{1jk} T_{ijk}$, while it is $(\beta_{0jk} + \beta_{2jk}) + (\beta_{1jk} + \beta_{3jk}) T_{ijk}$ in the treatment phase. β_{0jk} indicates the expected baseline level at the start of the treatment phase (when T=0), β_{1jk} the linear trend in the baseline scores. The coefficient β_{2jk} can then be interpreted as the immediate effect of the intervention on the outcome, whereas β_{3jk} gives an indication of the effect of the intervention on the trend.

The errors e_{ijk} are assumed to be distributed normally with the covariance matrix $\sigma^2\mathbf{I}$.
At the second level of the model, the variation over cases is described using four equations:

$$\beta_{0jk} = \theta_{00k} + u_{0jk} \qquad with \quad \mathbf{u} \sim N(0, \Omega_u)$$
$$\beta_{1jk} = \theta_{10k} + u_{1jk}$$
$$\beta_{2jk} = \theta_{20k} + u_{2jk}$$
$$\beta_{3jk} = \theta_{30k} + u_{3jk}$$

The first equation indicates that the baseline performance for subject *j* from study *k* equals an overall baseline performance for study *k*, plus a random deviation from this mean; the subsequent equations describe the variation over subjects from the same study of the time effect in the baseline condition, the immediate treatment effect, and the treatment effect on the linear trend, respectively.

At the third level, the variation of the study-specific regression coefficients from the second level equations is described:

$$\theta_{00k} = \gamma_{000} + v_{00k} \qquad with \quad \mathbf{v} \sim N(0, \Omega_v)$$
$$\theta_{10k} = \gamma_{100} + v_{10k}$$
$$\theta_{20k} = \gamma_{200} + v_{20k}$$
$$\theta_{30k} = \gamma_{300} + v_{30k}$$

Research Design:

Description of the research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

The value of the multilevel approach is evaluated by means of a simulation study. More specifically, we simulated data for MBD studies with starting points of the treatment phase staggered across the series.

Based on a study of Shadish and Sullivan (2011) of the characteristics of 809 single-case designs used to assess intervention effects, a survey of multiple baseline studies by (Ferron, Farmer, & Owens, 2010), and meta-analyses of SSED data (including, e.g., Alen, Grietens, & Van den Noortgate, 2009; Denis, Van den Noortgate, & Maes, 2011; Kokina & Kern, 2010; Shogren, Fagella-Luby, Bae, & Wehmeyer, 2004; Swanson & Sachse-Lee, 2000; Wang, Cui, & Parrila, 2011), we varied the following variables:

- the series length per case (*I*): 10, 20 and 40
- the number of cases per study (*J*): 3, 4 and 7
- the number of studies (*K*): 10 and 30
- the immediate treatment effect: 0 or 2
- the effect on the time trend: 0 or .2
- the between case variance: four diagonal elements of Ω_u: (2, .2, 2, .2), (.5, .05, .5, .05) and (8, .08, 8, .08),
- the between study variance: four diagonal elements of Ω_v: (2, .2, 2, .2), (.5, .05, .5, .05) and (8, .08, 8, .08),
- the within case variance: 1 and 5
- the between-study and between-case correlation in both kinds of effects: 0 and -.3

Crossing the levels of the seven factors leads to a 3x3x2x2x2x3x3x2x2 factorial design resulting in 2,592 simulation conditions. For each condition 2,000 data sets are simulated.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Data are simulated and analyzed in SAS, using the restricted maximum likelihood estimation procedure implemented in the procedure MIXED for multilevel or mixed models.

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)

The simulation study is currently being performed. Results of a pilot simulation study (with slightly other parameters) suggest that if data are not standardized, the estimates of the overall immediate treatment effect and the overall effect on the linear trend are almost unbiased. For standardized data, however, we find substantial positive bias in the estimates of the mean effects that strongly depends on the number of observations per case, as exemplified in Table 1.

(Please insert Table 1 here)

As expected, the variance in estimates was higher for standardized than for unstandardized data, because estimating the standardizing factor (the within case residual standard deviation) results in additional imprecision, especially when the number of observations per case is small. As a result of the bias in the mean estimate, the MSE is substantially higher for standardized data if the number of observations per case is small.

Standard errors were found relatively accurate for all conditions, but due to the bias the coverage proportion of the confidence intervals can decline to problematic levels when using standardized data. This is even more true with an increasing number of cases and/or studies, as illustrated in Table 2.

(Please insert Table 2 here)

More detailed results will be shown and discussed during the presentation.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

We conclude that the use of the multilevel model for combining single-case experimental data measured on the same scale yields accurate results, even with a small number of units at either level. Standardizing data within cases, however, has a fatal effect on the bias of the effect estimates and therefore on the confidence interval coverage proportions, unless the number of measurement occasions per case is large, say 50 observations per phase.

A limitation of the simulation study is that it only looks at the basic multilevel model, this is a model that does not account for autocorrelation, nonlinear trends, discrete dependent variables and so on. These extensions will be investigated in future research.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Alen, E., Grietens, H., & Van den Noortgate, W. (2009). Meta-analysis of single-case studies: An illustration for the treatment of anxiety disorders. Unpublished manuscript.

Denis, J., Van den Noortgate, W., & Maes, B. (2011b). Self-injurious behavior in people with profound intellectual disabilities: A meta-analysis of single-case studies. *Research in Developmental Disabilities, 32*, 911-923.

Center, B. A., Skiba, R. J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Special Education, 19,* 387-400.

Ferron, J. M., Farmer, J. L., Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel modeling approaches. *Behavior Research Methods*, *42*, 930-943.

Heyvaert, M., Maes, B., Van den Noortgate, W., Kuppens, S., & Onghena, P. (2011). A multilevel meta-analysis of single-case and small-n research on interventions for reducing challenging behavior in persons with intellectual disabilities. Manuscript submitted for publication.

Kokina, A., & Kern, L. (2010). Social story interventions for students with autism spectrum disorders: a meta-analysis. *Journal of Autism and Developmental Disorders, 40*, 812-826.

Nugent, W.R. (1996). Integrating Single-Case and Group-Comparison Designs for Evaluation. The Journal of Applied Behavioral Science, 32, 209-226. DOI: 10.1177/0021886396322007

Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation*, *113*, 95-109.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*. Retrieved from http://www.springerlink.com/content/f3r091184t30757j/fulltext.pdf

Shogren, K. A., Fagella-Luby, M. N., Bae, J. S., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior. *Journal of Positive Behavior Interventions, 6*, 228-237.

Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilities, 33*, 114-136.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*, 325-346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, *35*, 1-10.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*, 142-151.

Wang, S., Cui, Y., & Parrila, R. (2011). Examining the effectiveness of peer-mediated and video-modeling social skills interventions for children with autism spectrum disorders: a meta-analysis in single-case research using HLM. *Research in Autism Spectrum Disorders, 5*, 562-569.

Appendix B. Tables and Figures
Not included in page count.

Table 1: Bias in estimates of the mean immediate effect

		Unstandardized			**Standardized**		
J	*K*	*I*=10	30	50	*I*=10	30	50
2	10	.02	.01	.00	.47	.10	.05
	30	.00	.00	.00	.45	.08	.05
8	10	.00	.02	.00	.46	.11	.05
	30	.01	.01	-.01	.46	.10	.04

Table 2: Coverage proportion of the 90 % confidence intervals for the mean immediate effect

		Unstandardized			Standardized		
J	*K*	*I*=10	30	50	*I*=10	30	50
2	10	.91	.92	.89	.83	.91	.89
	30	.91	.91	.91	.55	.86	.88
8	10	.92	.92	.91	.72	.88	.89
	30	.91	.88	.89	.29	.86	.86