

The Application of Latent Curve Analysis to Testing Developmental Theories in Intervention Research¹

Patrick J. Curran²

University of North Carolina, Chapel Hill

Bengt O. Muthén

University of California, Los Angeles

The effectiveness of a prevention or intervention program has traditionally been assessed using time-specific comparisons of mean levels between the treatment and the control groups. However, many times the behavior targeted by the intervention is naturally developing over time, and the goal of the treatment is to alter this natural or normative developmental trajectory. Examining time-specific mean levels can be both limiting and potentially misleading when the behavior of interest is developing systematically over time. It is argued here that there are both theoretical and statistical advantages associated with recasting intervention treatment effects in terms of normative and altered developmental trajectories. The recently developed technique of latent curve (LC) analysis is reviewed and extended to a true experimental design setting in which subjects are randomly assigned to a treatment intervention or a control condition. LC models are applied to both artificially generated

¹This research was supported by Fellowship F32 AA05402 from the National Institute on Alcohol Abuse and Alcoholism, Grant 40859 from the National Institute on Mental Health, and a grant from OERI to the National Center for Research on Evaluation, Standards, and Student Testing. We thank Shepard Kellam for providing data for illustrations and Andrea Hussong, Siek-Toon Khoo, Jeff Shih, and Booil Jo for helpful comments and assistance. An earlier version of this paper was presented at the Workshop on a Scientific Structure for the Emerging Field of Prevention Research, cohosted by the Johns Hopkins University Prevention Research Center and the Prevention Research Branch, National Institute of Mental Health, December 5–6, 1994.

²All correspondence should be addressed to Patrick J. Curran, Department of Psychology, University of North Carolina—Chapel Hill, Chapel Hill, North Carolina 27599-3270.

and real intervention data sets to evaluate the efficacy of an intervention program. Not only do the LC models provide a more comprehensive understanding of the treatment and control group developmental processes compared to more traditional fixed-effects models, but LC models have greater statistical power to detect a given treatment effect. Finally, the LC models are modified to allow for the computation of specific power estimates under a variety of conditions and assumptions that can provide much needed information for the planning and design of more powerful but cost-efficient intervention programs for the future.

KEY WORDS: latent curve analysis; developmental theories; intervention research.

Assessing the efficacy of an intervention program can often be an imposing task. Issues such as longitudinal data, nonequivalent control groups, subject attrition, and non-normally distributed measures can all pose challenges when trying to make inferences about the utility of an intervention program. Another complexity commonly encountered in prevention and intervention research is the need to consider development and individual differences in development over time. Traditionally, the effectiveness of an intervention program has been assessed in terms of the group level mean and variance of the targeted behavior of interest. True random assignment to condition attempts to equate the treatment and control groups prior to the implementation of the intervention (e.g., Cook & Campbell, 1979), and the treatment effect is typically measured as the difference between the within-group means and the within-group variances on the outcome behavior at the end of the intervention trial. Examples of commonly used analytic techniques include the *t* test, ANOVA, ANCOVA, MANOVA, MANCOVA, multiple regression, and fixed effects structural equation modeling. Although these analytic approaches can be quite useful for evaluating treatment effects under certain rather restrictive assumptions, there are several potential limitations that arise when these techniques are used to study development over time. Examples include decreased statistical power, potentially biased parameter estimates, inability to model individual differences in change, and unnecessary restriction of inferences that can be obtained from the observed data (Muthén & Curran, 1997; Rogosa, 1988; Rogosa & Willett, 1985; Willett, 1991). These limitations can be particularly salient when attempting to assess the degree to which a prevention or intervention program influences developmental trajectories over time.

Traditional Longitudinal Data Analytic Techniques

There is a multitude of analytic approaches that have been developed over the years to evaluate longitudinal data. One of the earliest longitudinal

analytic techniques was to study the *raw change* score. Change was computed as the difference between the Time 1 and the Time 2 scores, and these raw change scores were examined as a function of individual or group characteristics. Although the use of raw change scores has been heavily criticized (e.g., Cronbach & Furby, 1970), many of these early criticisms were later determined to be unfounded (Rogosa, 1988). Raw change scores are typically analyzed using approaches such as *t* tests, analysis of variance (ANOVA) models, and multiple regression models. An alternative approach to studying the raw change score is the *residualized* change score. Here, change is computed as the error (or *residual*) between the observed Time 2 score and the expected Time 2 score as predicted by the Time 1 score. Residualized change scores are typically analyzed using multiple regression models (with the Time 1 measure as a predictor and the Time 2 measure as the criterion) or ANCOVA models (with the Time 1 measure as a covariate and the Time 2 measure as the criterion). Although widely used in behavioral research, the study of residualized change scores has been criticized on both statistical and theoretical grounds (Rogosa, 1987, 1988; Rogosa & Willett, 1985; Stoolmiller, Duncan, Bank, & Patterson, 1993; Willett, 1991), and it is usually recommended that these be avoided when possible. Both raw and residualized change scores consider only change between two discrete time points, although additional time points can be incorporated [as in the autoregressive cross-lagged panel design (Dwyer, 1983)]. These multiple-time point models are typically a compiled series of two time point comparisons and are thus only a modest extension of the residual change model.

Despite several advantages of these techniques, a problem quickly arises when using these approaches to evaluate developmental intervention theories. In many intervention trials, particularly those involving children or adolescents, the interest is not so much in the absolute level of a targeted behavior at a *particular* time point (e.g., the mean level immediately following the intervention) as it is in the developmental trajectory of the targeted behavior across *multiple* time points (e.g., the continuous developmental process before, during, and after the intervention). Frequently, behaviors that are the focus of intervention programs show natural systematic development over time. Examples include reading comprehension, alcohol use, aggressive behavior, juvenile delinquency, and cognitive reasoning ability. The goal of the intervention is to alter the normative developmental trajectory of the targeted behavior over time. An intervention might be designed to accelerate the normative developmental process of an adaptive behavior (e.g., an intervention designed to increase a child's reading and comprehension skills) or, alternatively, to decelerate the normative developmental process of a maladaptive behavior (e.g., an intervention designed to slow

a child's escalation of alcohol use). When recasting treatment evaluation in terms of development, the effectiveness of an intervention is the degree to which the intervention can alter the normative developmental trajectory that exists without exposure to the treatment.

The traditional analytic technique often times best suited for studying change across multiple time points is the repeated-measures MANOVA model, which is sometimes referred to as *trend analysis*. In this approach, change is construed as a linear (or higher-order) trend over the multiple time points, and group differences are examined with respect to the magnitude of the trend relative to the variance within each group. For example, say that five repeated-measures are assessed on a given construct over time. A linear trend is fit to the five repeated measures, and differences in the magnitude of the trend are assessed as a function of categorical independent grouping variables. If the repeated measures of a single construct are considered, this model is sometimes referred to as a *singly multivariate* MANOVA. If more than one construct is considered, this model is referred to as a *doubly-multivariate* MANOVA. Finally, if a continuous covariate is included, this model is referred to as a repeated-measures MANCOVA.

Advantages of the repeated measures MANOVA model include the ability to model group trajectories over time and the ability to adjust for differences in trajectories as a function of continuous covariates. However, this approach is often severely limited in that only the effects of categorical predictor variables can be explicitly modeled (just as in a basic ANOVA model), and more importantly, individual differences in rates of change over time within group are attributable to error (also just as in a basic ANOVA). Because of these (and a number of other) limitations, the repeated-measures MANOVA model is often not an ideal analytic technique for studying predictors and correlates of individual differences in development over time. It is thus important that new analytic methods be developed and disseminated that are better suited to address these complex questions so that stronger and more informed inferences can be made about the efficacy of developmental prevention and intervention programs.

Latent Curve Analysis

One class of analytic techniques that are designed better to address questions of individual differences in change over time are broadly referred to as *random coefficients* models. Variations of these models can be found in biometrics (Rao, 1958; Laird & Ware, 1982; Diggle, Liang, & Zeger, 1994), education (Cronbach, 1976; Burstein, 1980; Bock, 1989; Bryk & Raudenbush, 1992), and psychometrics (McArdle, 1988; McArdle & Ep-

stein, 1987; Meredith & Tisak, 1984, 1990; Tucker, 1958). The psychometric tradition defines individual differences in growth trajectories in terms of latent variables which thus fits nicely within the structural equation modeling framework (Bentler, 1980; Jöreskog & Sörbom, 1979; Sörbom, 1982).

A general overview of latent curve (LC) models is presented here. More detailed information is given by Browne and Du Toit (1991), McArdle (1986, 1988, 1989, 1991), McArdle and Epstein (1987), Meredith and Tisak (1984, 1990), Muthén (1991, 1994), Muthén and Curran (1997), and Willett and Sayer (1994). LC models are fit to the observed covariance matrix and vector of means using any structural modeling software program (e.g., Amos, CALIS, EQS, Mplus, LISREL, Mx). The basic LC model is comprised of two latent factors. The first latent factor represents the initial status (or intercept) and is defined by fixing all of the factor loadings (or bases) of the repeated measures to 1.0. This factor captures the starting point of the developmental growth trajectory at Time 1. The second latent factor represents the growth rate over time (or slope), and the factor loadings of the repeated measures are a mixture of fixed or freely estimated parameters that define the shape of the developmental growth trajectory over time. The means of the initial status and growth rate factors represent the group parameter values of the intercept and slope of the developmental trajectory. The variances of the initial status and growth rate factors represent the individual variability of each subject around the group parameters. Larger factor variances reflect greater individual differences in growth over time whereas smaller factor variances reflect more similar patterns of growth over time. Finally, the variance in these growth factors can be modeled as a function of additional explanatory variables (e.g., age, gender, treatment condition, temperamental characteristics) to understand better observed individual variability in rates of change over time.

A powerful application of this basic LC model is in the examination of treatment effects within an experimental design. McArdle (1989) first demonstrated the advantages of this approach with his multiple-group application of LC models within both true and quasi-experimental design settings. Muthén and Curran (1997) proposed an extension of this LC model which allows for the explicit modeling of individual variability in growth trajectories associated with a treatment intervention program implemented using a true experimental design. What follows is a brief summary of this procedure. See Muthén and Curran (1997) for further details.

The first step in this LC modeling process is to estimate the normative developmental trajectory of the targeted behavior within just the control group. This allows for the definition of the specific level, shape, and variability of the normative developmental trajectory of the targeted behavior as it exists without exposure to the treatment intervention. Next, the treatment

group is added as a second independent subgroup using traditional multiple group analysis in structural equation modeling (e.g., Jöreskog, 1971; Sörbom, 1982). The parameters describing the normative developmental trajectory previously identified in the control group are also estimated in the treatment group, and these parameters are equated across the two groups. This equating allows for the identification of the portion of growth in the treatment group that is attributable to the normative developmental process observed within the control group. However, in addition to the normative growth factors, a second growth rate factor is added that is unique to the treatment group (resulting in a total of three growth factors). This added factor allows for the identification of *differential* growth that exists within the treatment group above and beyond the normative developmental trajectory that exists within the control group. This factor captures the degree to which the normative developmental trajectory observed in the control group was altered as a function of the intervention applied to the treatment group. It is this unique treatment group growth factor that explicitly captures the intervention treatment effect.

Advantages of LC Models for Testing Developmental Intervention Theory

Defining intervention effects in terms of altered and unaltered developmental trajectories offers a number of key advantages to the intervention researcher. First, this formulation avoids many of the limitations associated with evaluating treatment effects and change over time using the more traditional fixed-effects models (Rogosa, 1987, 1988; Rogosa & Willett, 1985; Willett, 1991). Second, many times our interest is not only in the mean differences between groups (as tested by the regression-based models), but also in group differences in variances and covariances, information that is often lost in more traditional types of fixed effects models (McArdle, 1989). Third, and possibly most important, discrete time-specific assessments are generally not consistent with the underlying tenets of psychosocial theories of development and change. Developmental theories typically do not posit change in terms of time-specific comparisons (e.g., females are expected to be higher than males on a certain attribute at Time 2 above and beyond their previous level of standing relative to the group mean at Time 1). Instead, developmental theory tends to construe change as a *continuous* growth process over time, and this process is described in terms such as individual differences in onset, escalation, acceleration, plateau, and deceleration. Modeling time-specific comparisons typically cannot capture these complex types of relations over time. LC models not only are better suited

statistically to model change over time, but also are more consistent with the basic formulations of developmental theory. This increased consistency allows for a stronger test of the theoretically derived hypotheses, which in turn increases the level of understanding about the developmental theories in general.³

LC models provide several other advantages to the intervention scientist as well. For example, it is known that not all interventions work for all people (Kellam *et al.*, 1991). It is thus important to identify those subjects who most benefit from the intervention. Because LC models allow for the estimation of individual differences in change over time, differential response to treatment can be examined in an attempt to identify factors associated with stronger or weaker responsiveness to treatment [similar to an aptitude-treatment interaction (Cronbach & Snow, 1977)]. Another advantage is that not only do LC models have greater power to detect a treatment effect compared to more traditional fixed effect models (Muthén & Curran, 1997), but LC models can be readily modified to allow for the computation of specific estimates of power under a variety of assumptions and conditions. Using techniques described by Satorra and Saris (1985), characteristics such as the shape and rate of normative development, treatment effect size, number of repeated measures, and sample size can be defined within both the control and the treatment groups. Estimates can then be obtained about the minimum sample size necessary to detect a given treatment effect at a given level of power. Examples of questions that can be examined using this LC technique include the costs and benefits of utilizing a balanced design (with equal numbers of treatment and control subjects) versus the effects of adding more control than treatment subjects (because control subjects may be less expensive), understanding how long to follow the experimental groups during and after the intervention, and understanding the advantages and disadvantages of adding more subjects with fewer repeated observations versus fewer subjects and more repeated observations.

The Present Paper

Muthén and Curran (1997) presented a technical description of a new LC methodology that they proposed for examining individual differences in change over time within a broad class of experimental design settings. The present paper attempts to focus these techniques in a less technical fashion specifically on the evaluation of developmental preventive interven-

³For a general discussion of these and other related issues, see Wohlwill (1991) and the subsequent commentaries.

tion programs with children and adolescents. To this end, LC models are first applied to artificial data to demonstrate basic model building techniques and methods for power estimation. This is followed by an application of LC models to an actual intervention data set to demonstrate real-world model-fitting procedures and interpretation of results.

METHOD

Description of Data Sets

Artificial Data Set: Linear Treatment Effect with Interaction Between Treatment and Initial Status

The artificial data set was designed to reflect a hypothetical intervention study in which there was successful randomization to condition at Time 1, the intervention was implemented immediately following the Time 1 measure, and there was a total of five equally spaced repeated observations of the behavior of interest. For the control group, the intercept and slope values corresponded to a normative growth rate of one standard deviation over the five time points (the intercept of the normative growth trajectory was set to $\mu = 1.0$ with a variance of $\sigma^2 = 1.0$, the slope over the five time periods was set to $\beta = .798$ with a variance of $\sigma^2 = .20$, and a correlation of $\rho = .25$ was defined between the intercept and the slope factors). For the treatment group, the intercept and slope values corresponded to a growth rate that was 23% higher in the treatment group compared to the control group (the initial starting point of the growth trajectory for the treatment group was set to $\mu = 1.0$ with a variance of $\sigma^2 = 1.0$, the rate of growth was set to $\beta = .981$ with a variance of $\sigma^2 = .20$, and a correlation of $\rho = .25$ was defined between the intercept and the slope factors). These values were chosen so that the difference between the treatment and the control group means at Time 5 scaled by the pooled standard deviation at Time 5 represented a small effect size ($d = .20$) in Cohen's (1988) terms. The residuals of the repeated measures were specified to have equal proportions of error variance across time (e.g., residual $R^2 = .50$ for all measures) and the residuals were uncorrelated across time. Finally, there was an interaction between the treatment condition and the initial status such that subjects with a higher initial status benefited more from the intervention than subjects with a lower initial status. This interaction was captured in a nonzero regression parameter ($\gamma = .252$) between the intercept and the treatment slope factors estimated within the treatment group. This regression parameter was chosen to represent a moderately small effect

size ($ES = .30$).⁴ The total sample size used for the artificial data set was set to be $N = 500$, with $N = 250$ cases in the control group and $N = 250$ cases in the treatment group. The covariance matrices and mean vectors for the treatment and control groups are presented in Table I.⁵

Real Intervention Data: Eight-Time Point Baltimore Aggressive Behavior Intervention

Overview. The intervention data set was drawn from a large developmental epidemiologically based preventive trial implemented in the Baltimore City Public School System. The details of the study design and recruitment are presented by Kellam, Rebok, Ialongo, and Mayer (1994). Of interest to the present paper are the results from the Good Behavior Game (GBG) intervention and matched control (CON). The GBG is a 2-year long classroom-based behavior management program designed to promote positive behaviors and decrease disruptive and aggressive behaviors. A total of eight repeated measures was obtained on all subjects. The first four measures were assessed during the fall and spring semester of the 2 years of the intervention. The next four measures were assessed only during the spring of the 4 years following the end of the intervention.

Table I. Covariance Matrices and Mean Vectors for the Artificial Data Set

Control group					
Time 1	2.0000000				
Time 2	1.1118034	2.8472136			
Time 3	1.2236068	1.7354102	4.4944272		
Time 4	1.3354102	2.0472136	2.7590170	6.9416408	
Time 5	1.4472136	2.3590170	3.2708204	4.1826238	10.188854
Mean	1.0000000	1.7979980	2.5959960	3.3939940	4.1919920
Treatment group					
Time 1	2.0000000				
Time 2	1.4207534	4.6211804			
Time 3	1.8415068	3.2004270	9.1186946		
Time 4	2.2622602	4.0902639	5.9182675	15.4925420	
Time 5	2.6830136	4.9801007	7.2771878	9.5742749	23.742724
Mean	1.0000000	1.9809430	2.9618860	3.9428290	4.9237720

⁴The effect size for the interaction was not cast in Cohen's (1988) terms but, instead, was computed as a function of the latent growth factors (for further details see Muthén & Curran, 1997).

⁵The raw data and corresponding computer code used in all of the following analyses can be obtained from the first author or can be downloaded via the Internet from <http://www.unc.edu/~curran>.

Subjects. The full sample consisted of $N = 1084$ children enrolled in the first grade in the Baltimore City Public School System. Of these, $N = 693$ children remained in the same intervention or control condition for the first 2 years, and $N = 590$ were available at the final year of measure. The present paper considered only children in the GBG for both years of the intervention and the matched control group. Based on Kellam and co-workers (1994) findings that the effectiveness of the GBG was limited to males, we also considered only males in the GBG and CON conditions. Finally, because a large proportion of the data was missing for at least one of the eight time points (approximately 55%), the missing values were imputed using the mean of the immediately adjacent nonmissing values for each individual subject.⁶ The final sample sizes considered for the present paper were $N = 75$ for the GBG condition and $N = 111$ for the CON.

Measures. Aggressive behavior was assessed using the Teacher Observation of Classroom Adaptation—Revised (TOCA-R). The TOCA-R measures the frequency of 18 types of aggressive behavior using a 6-point response scale that ranged from almost never to almost always. TOCA-R measures were obtained during the fall and spring of the first 2 years (the period of the intervention) and just in the spring for the following 4 years. This resulted in a total of eight repeated measures covering a 6-year period, 2 years of intervention and 4 years postintervention. A single TOCA-R score was obtained for each child at each time period. The covariance matrices and mean vectors for the treatment and control groups are presented in Table II.

RESULTS

Overview of LC Modeling Strategy

The proposed modeling strategy involves five basic steps.⁷ The first step is to model the developmental process within just the control group. This allows for the identification of the level, shape, and variability of the natural developmental process as it exists without exposure to the intervention. The second step is to model the developmental process within

⁶It is important to note that, because of the imputation techniques used here, these data are not necessarily reflective of the complete data set of Kellam *et al.* (1994). We do not recommend this type of data imputation in general, but this was done to simplify the demonstration of the proposed techniques.

⁷Many additional steps could be incorporated in this model building process, the specific selection of which would depend upon the particular data and hypotheses at hand. See McArdle (1989, 1991) for a detailed discussion and application of many useful alternative model building techniques.

Table II. Covariance Matrices and Mean Vectors for the Baltimore Intervention Data Set

Control group									
Time 1	1.026								
Time 2	0.852	0.002							
Time 3	0.632	0.697	1.111						
Time 4	0.572	0.561	0.959	1.226					
Time 5	0.554	0.609	0.783	0.885	1.321				
Time 6	0.613	0.601	0.690	0.713	0.811	1.251			
Time 7	0.680	0.690	0.792	0.791	0.907	1.007	1.353		
Time 8	0.530	0.539	0.779	0.852	0.947	0.967	1.021	1.490	
Mean	1.952	2.061	1.984	2.190	2.307	2.213	2.318	2.256	
Treatment group									
Time 1	1.654								
Time 2	1.076	1.340							
Time 3	0.982	0.821	1.274						
Time 4	0.870	0.673	1.034	1.385					
Time 5	0.870	0.873	0.658	0.781	1.397				
Time 6	0.821	0.794	0.551	0.695	0.958	1.464			
Time 7	0.481	0.503	0.548	0.563	0.461	0.585	0.817		
Time 8	0.364	0.582	0.452	0.521	0.491	0.613	0.604	1.344	
Mean	2.307	2.211	2.268	2.356	2.489	2.451	2.304	2.264	

just the treatment group. Although the presence of treatment effects cannot be formally assessed without the simultaneous consideration of the control group, it is useful to examine the characteristics of the developmental trajectory observed within the treatment group prior to moving to the multiple-group analyses. The third step is to model the developmental processes within the control and treatment groups simultaneously. This allows for the modeling of the portion of growth in the treatment group that is attributable to the normative developmental process, and any remaining difference in the treatment group growth can be attributable to the intervention. The fourth step is to test for the potential interaction between initial status and the magnitude of the treatment effect. The final step is to perform a sensitivity analysis to assess the relative impact of the various assumptions and restrictions imposed on the model. The first four steps are discussed in detail below. See Muthén and Curran (1997) for a discussion of the fifth step.

Artificial Data

Development in the Control Group

The first step was to fit a LC model within just the control group. A two-factor, five-indicator LC model was estimated based on the artificially

generated covariance matrix and mean vector using LISCOMP (Muthén, 1987).⁸ The loadings on the intercept factor were all fixed to 1.0 to reflect the initial status, and the loadings on the slope factor were set to 0, 1, 2, 3, and 4 to reflect the equally spaced observations and linear growth over time (see Fig. 1A). Note that, although not shown in the figure, both the variance and the mean are estimated for each latent factor [see McArdle (1988) for a diagramming approach that explicitly incorporates these mean and variance estimates]. The correlation between the intercept and the slope factors was freely estimated, as well as the variances of the repeated measures.⁹ Because the model estimated in the sample corresponded precisely to the model that existed in the population (due to the artificially generated data), the estimated model fit the data perfectly [$\chi^2(10, N = 250) = 0, p = 1.0$]. There was a significant ($p < .05$) mean estimate for the linear slope factor that reflected a group-level linear developmental trajectory over time. There were significant variance estimates in both the intercept and the slope factors that reflected the presence of important individual differences in the initial status and change over time. Thus, there were both meaningful development and individual differences in development within the control group, and this represented the normative developmental process as it existed without exposure to the intervention.

Development in the Treatment Group

The next step was to repeat the LC modeling within just the treatment group. The same modeling techniques were applied to the treatment group as described for the control group. This model also fit the observed data perfectly [$\chi^2(10, N = 250) = 0, p = 1.0$]. The results closely matched those of the control group. There was a significant mean estimate for the linear slope factor, and there were significant variance estimates for both the intercept and slope factors. This suggests that both groups were characterized by linear growth over time and that there was important individual variability in both starting point and change over time.

⁸Although LISCOMP was used to estimate the models presented here, all of the analyses could be equivalently estimated using any structural equation modeling software package that can examine latent variable covariance and mean structures (e.g., Amos, CALIS, COSAN, EQS, LISREL, Mx, etc).

⁹This correlation could have been estimated as a regression parameter instead, and all of the resulting conclusions for the present analyses would have remained the same. However, there may be applications in which a regression parameter might better model the relation between the intercept and the slope factors. See Curran, Stice, and Chassin (1997) for an example of this type.

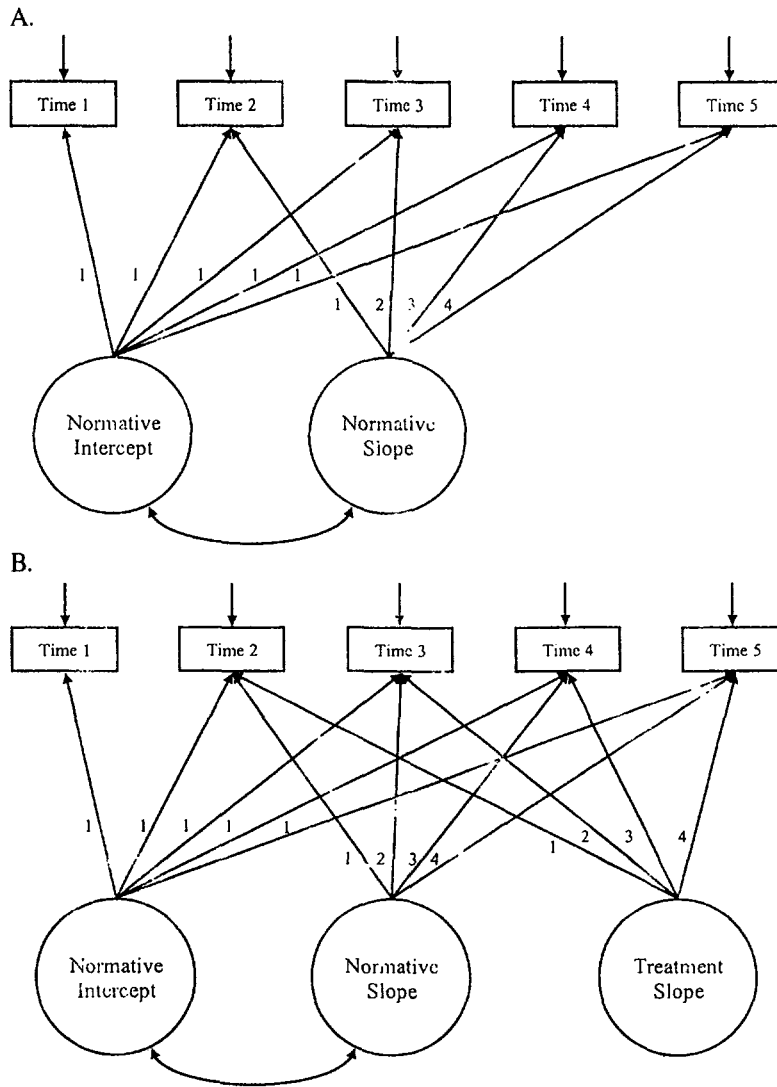


Fig. 1. Path diagram of a five-time point latent curve model within the control group (A) and within the treatment group (B).

Multiple-Group LC Model

The third step of the modeling procedure was to estimate the LC model simultaneously for the control and treatment groups using the multiple-group estimation procedure of LISCOMP. The normative growth factors previously identified in the control group were estimated within the treatment group, and the factor means, variances, and covariances were equated across the two groups. These equality constraints identified the portion of growth in the treatment group that was attributable to natural development over time. Although a lack of equality in the growth factors between the treatment and the control group would suggest the existence of a treatment effect, freely estimating the mean and variance of the growth factors within each group would still confound the variability associated with normative development and treatment-affected development occurring within the treatment group. To avoid this potential confound and explicitly model that part of growth that might be directly attributable to the treatment intervention, a second linear growth factor unique to the treatment group was added, and the mean and variance of this treatment growth factor were freely estimated. This factor was designed to capture any treatment effect that existed over and above the normative developmental process observed in the control group. Thus, in Fig. 1, the upper model (Fig. 1A) was estimated in the control group, whereas the lower model (Fig. 1B) was estimated in the treatment group. This two-group model was estimated and fit the data perfectly [$\chi^2(23, N = 500) = 0, p = 1.0$]. Of most importance were the significant mean and variance estimates of the treatment growth factor. The significant positive mean estimate reflected a meaningful treatment effect over and above the normative developmental trajectory of the control group. The significant variance estimate reflected the existence of meaningful differences in the effectiveness of the treatment across individuals. Based on the significant mean and variance estimates of the treatment growth factor, it was concluded that the hypothetical treatment intervention resulted in a significant linear acceleration of the normative developmental trajectory that existed without exposure to the treatment.

Estimation of the Interaction Effect

Given that there was a significant treatment effect, one key question to be explored was whether there was an interaction between initial status and change over time. That is, did the magnitude of the treatment effect vary as a function of initial status? To test for the potential interaction

effect between initial status and change over time, the treatment slope factor was regressed on the intercept factor within the treatment group. This regression parameter was positive and significant, reflecting that the individual response to the treatment condition varied as a function of initial status. Plotting the model implied treatment growth trajectories as a function of initial status reflected that subjects with a higher initial status tended to grow at significantly steeper rates over time (and thus benefited more from the treatment) compared to subjects with a lower initial status. Thus, although there was an overall main effect of treatment, there was a stronger intervention effect for a certain subset of individuals. This is important information that is needed for the proper evaluation of the efficacy of the intervention.

Power Estimation

The information obtained from the above models can now be used to compute specific estimates of effect size and power that can provide important information for the design of future studies. There are two power estimates of interest here: the power to detect a given main effect of treatment and the power to detect a given interaction between the treatment effect and initial status. The power estimates were computed using the procedures presented by Satorra and Saris (1985) and applied to the LC modeling framework of Muthén and Curran (1997). The first step when estimating power is to compute the model implied covariance matrix and mean vector for both the control and the treatment groups as they are hypothesized to exist within the population [see Muthén & Curran (1997) for the technical details of this procedure]. These matrices can reflect any shape, rate, and variability in development for both the treatment and the control groups. Once computed, the LC model is then fit to the population matrices and, when defined to correspond exactly to the population model, should result in a model chi-square equal to zero (as was obtained in the example above). Next, the mean of the treatment growth factor is fixed to zero within the treatment group, and the model is reestimated. The resulting nonzero model chi-square value provides an estimate of the noncentrality parameter, or λ . This value, when cross-referenced in a noncentral chi-square table under 1 df (because only one parameter was fixed to zero) provides the estimate of statistical power that exists to detect the misspecification under the given sample size. Noncentral chi-square tables are presented by Saris and Stronkhorst (1984) or can be computed using the PROBCHI function in SAS (SAS Institute, Inc., 1994) or any other basic statistical program.

The effect sizes for the power analysis were computed using Cohen's (1988) measure d for mean differences. The effect size reflected the difference between the treatment and the control means at Time 5 divided by the pooled variance at Time 5. It is recommended that a power of at least .80 be attained for social science research (Cohen, 1988). For comparative purposes, we first estimated the power to detect three effects ranging from small to moderate in size ($d = .20, .30, \text{ and } .50$) using a traditional ANCOVA model examining differences between the Time 5 scores after having adjusted for Time 1 levels (see Fig. 2).¹⁰ The power curves reflect that for the ANCOVA model, a sample size of $N = 725$ is required to obtain a power of .80 to detect a treatment effect size of $d = .20$, $N = 325$, for $d = .30$, $N = 120$, for $d = .50$. Power estimates were then computed to detect the same effect sizes using the LC model based on all five time points (i.e., the two-group model presented above) using the Satorra-Saris method described earlier (see Fig. 2). The LC models were found to have substantially greater power to detect the very same effect compared to the power that was estimated for the ANCOVA models. Using the LC design, a power of .80 is obtained at $N = 525$ for $d = .20$, $N = 225$ for $d = .30$, and $N = 80$ for $d = .50$. Thus, the LC model requires nearly 30% fewer subjects to detect a small effect size at a power of .80 than is required for the same effect size and same power for the ANCOVA model.

These results can readily be expanded in many useful ways. For example, say that a treatment effect size of $d = .20$ is expected at Time 5, but a researcher would like to calculate the number of subjects required for a LC model to maintain a power of .80 as a function of the total length of time of the study. As shown above, given a $d = .20$ at Time 5, $N = 525$ subjects are required for a power of .80 if the sample is followed for all five time periods using the LC model. If only the first four time periods are considered, the sample must be increased to $N = 650$ to maintain a power of .80, and if only the first three time periods are considered, the sample must be increased to $N = 950$. If the first, third, and fifth time periods are considered, $N = 600$ cases are required for a power of .80, and if the growth process were to be followed for seven time periods, $N = 425$ cases would be required to maintain a power of .80. So, if only three time periods are to be considered, 37% fewer subjects would be required to detect a treatment effect if the three measures are spread over the 5 years compared to if the three measures are taken during the first 3 years.

¹⁰We could have similarly estimated power for the repeated-measures MANOVA model, but we chose to present comparative power estimates for the ANCOVA model because this approach tends to be more commonly used in applied research settings compared to the MANOVA approach. See McArdle (1989) for the LC equivalent to the repeated-measures MANOVA model.

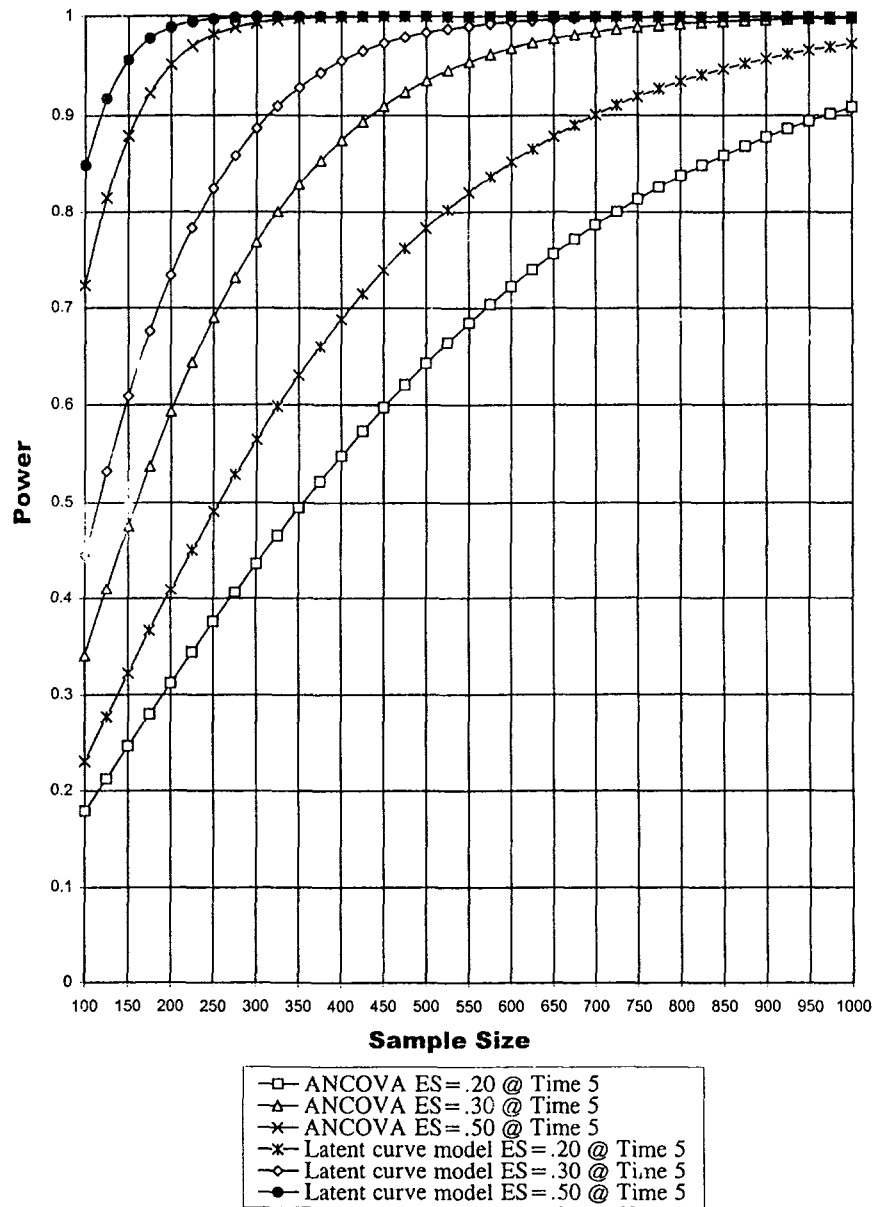


Fig. 2. Power to detect a main effect of treatment as a function of effect size and sample size for the latent curve model and the ANCOVA model.

Power can also be estimated to detect an interaction effect between initial status and change over time. All of the power estimates for the interaction effect assume a small ($d = .20$) main effect of treatment at Time 5. Figure 3 presents power curves for five effect sizes. The computation of these effect sizes differs from those discussed by Cohen (1988) because the effects are expressed using the latent growth factors and not just in terms of the observed variables. The effect sizes considered here are thought to represent small to moderate interaction effects. It can be seen that there are large increases in power when moving from the smallest to larger effect sizes, but the increase in power diminishes as the effect sizes increase. It is also evident that it is more difficult to detect an interaction effect between initial status and change over time compared to detecting a main effect of treatment. This is consistent with findings in multiple regression models (e.g., Aiken & West, 1991). Figure 3 shows that for the smallest effect size, even $N = 1000$ subjects are not sufficient to detect the interaction at a power of .80. However, the small increment in effect size from .20 to .25 results in a decrease in the sample size requirement from $N = 1000$ to $N = 700$ to maintain a power of .80, and an effect size of .30 requires a further reduction in sample size to $N = 500$.

It is clear, then, that two power estimates must be considered. The previous analyses indicated that $N = 525$ subjects would be required to detect a small main effect of treatment over five time periods. However, in the presence of a small main effect of treatment, $N = 700$ subjects would be required to detect a small interaction between initial status and change over time. Power estimates for both the main effect of treatment and the interaction effect of treatment and initial status must be considered when estimating the necessary number of subjects for future studies.

Real Intervention Data: Baltimore Aggressive Behavior Intervention

Development in the Control Group

A two-factor eight-indicator LC model was fit to the $N = 111$ subjects in the control group. The factor loadings on the growth factor were set to be 0, 1, 2, 3, 5, 7, 9, and 11. The one-unit increments of the first four loadings represented the 6-month interval between the first four measures, whereas the two-unit increments of the second four loadings represented the 12-month interval between the second four measures. This model fit the observed data well [$\chi^2(24, N = 111) = 31.9, p = .13$]. There was a significant positive linear component of the developmental trajectory as well as significant individual variability in both initial status and change over time. Next,

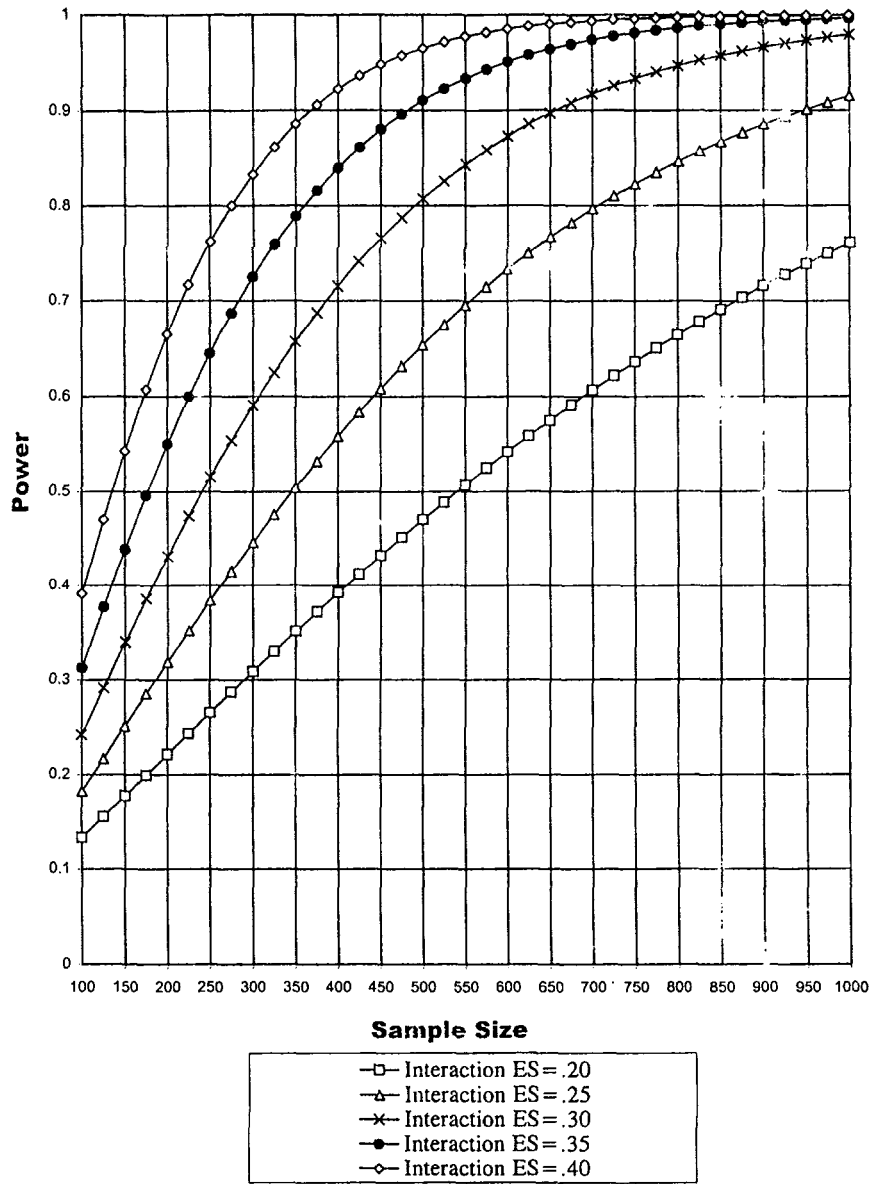


Fig. 3. Power to detect an interaction between treatment and initial status as a function of effect size and sample size for the latent curve model.

because examination of the raw data suggested the possibility of a quadratic component of growth (see mean vectors in Table II), this linear model was reestimated with a second added growth factor with factor loadings of 0, 1, 4, 9, 25, 49, 81, and 121. These represented the squared linear loadings and were designed to capture the potential quadratic component of growth (see, e.g., Bryk & Raudenbush, 1992).^{11,12} The mean and variance of the quadratic factor were estimated, as well as the covariances among the growth factors. Estimation problems were encountered because of a zero estimate in the variance of the quadratic factor, so this variance was set to zero [to represent a fixed quadratic effect (Bryk & Raudenbush, 1992)]. Note that because the quadratic factor was a fixed effect and thus had no variance, this factor was not correlated with the remaining growth factors. This model converged to a proper solution and fit the data well [$\chi^2(23, N = 111) = 24.9, p = .35$]. Comparing the models with and without the quadratic factor resulted in a significant chi-square difference test [$\chi^2(1, N = 111) = 6.9, p = .01$], reflecting that the addition of the quadratic factor resulted in a significant improvement in model fit.¹³ In summary, the normative developmental process within the control group was composed of a significant positive linear random component and a significant negative quadratic fixed component, thus reflecting a general increase in aggression over time with a slight downturn at the end of the growth trajectory.

Development in the Treatment Group

The same linear model described above was fit to the eight repeated measures in the $N = 75$ treatment subjects, and this model fit the data well [$\chi^2(24, N = 75) = 32.9, p = .11$]. A quadratic factor was then added to the model, and as found in the control group, estimation problems were encountered given a zero variance estimate in the quadratic factor. This variance was fixed to zero to reflect a fixed quadratic effect, and this model also fit the data well [$\chi^2(23, N = 75) = 30.6, p = .14$]. Although the chi-

¹¹Note that issues of multicollinearity associated with failing to center quadratic effects in multiple regression models (e.g., Aiken & West, 1991) do not apply here because the factor loadings in the LC models are not predictor variables in the equation but, instead, are model coefficients that convey the impact of the latent factor scores on the observed repeated measures. The loadings in these models could have been centered, but this would change the subsequent interpretation of the parameter estimates. See Bryk and Raudenbush (1992) for further discussion about these scaling issues in growth models.

¹²There are a number of approaches available to estimate nonlinear effects in growth models, but space constraints preclude us from fully exploring these options here. See McArdle (1989, 1991) for a further discussion of these issues.

¹³It can be shown that the difference between two chi-square test statistics is itself distributed as a central chi-square and can thus be used to test whether added restrictions to a model resulted in a significant decrement in overall model fit (see, e.g., Bollen, 1989, p. 400).

square difference test [$\chi^2(1, N = 75) = 2.2, p = .13$] between the two models was nonsignificant, these results suggest that the observed developmental processes were of similar functional form for both the treatment and the control groups. A formal comparison of the growth processes between the treatment and control groups is tested in the multiple-group analysis.

Multiple-Group LC Model

The models described above were then estimated simultaneously for the treatment and control groups so that both groups were identified by the same normative growth factor structure, but an additional fixed-effects linear growth factor was estimated that was unique to the treatment group. This linear treatment growth factor was designed to capture explicitly any added downturn in the normative developmental trajectory of aggressive behavior attributable to the GBG. This model was estimated and fit the data moderately well [$\chi^2(51, N = 186) = 74.9, p = .02$]. The parameter estimate of key interest was the nonsignificant ($p > .20$) mean estimate of the linear treatment growth factor. The substantive conclusion at this point was that the intervention did not significantly alter the natural developmental trajectory that existed without exposure to the intervention. However, the potential interaction effect must first be considered prior to accepting this conclusion.

Examination of the Interaction Effect

The multiple-group model was reestimated as described above with one exception: the linear treatment growth factor was regressed upon the intercept factor. This model fit the data well [$\chi^2(50, N = 186) = 64.5, p = .08$]. There were two important findings of interest. First, the intercept of the linear treatment growth factor remained nonsignificant. This reflected that there was no main effect of the treatment over time. However, there was a strong and significant ($p < .001$) negative regression parameter estimate between the intercept and the linear treatment growth factor. This reflected that there was indeed a treatment effect, but the effect varied in magnitude as a function of initial status. Probing of this interaction effect revealed that treatment subjects with a higher initial status tended to decrease in aggressive behavior more rapidly toward the end of the eight time periods compared to those subjects with a lower initial status. Note that had the interaction effect not been considered, the conclusion would have been that the intervention was not efficacious. However, inclusion of

information about initial status revealed a strong and consistent treatment effect, but only for a certain subset of individuals.

Power Estimation

There are two relevant questions regarding power in this example. The first relates to assessing the actual obtained effects sizes and level of power for the above analyses. The second incorporates information about these findings that might better inform intervention studies similar to this in the future. Regarding the first question, consider the hypothetical situation in which the previous analyses revealed no evidence for the existence of a treatment effect. There are two possible interpretations of this null finding. First, the intervention may simply not have been successful, and the normative developmental trajectory was not altered as a function of the intervention. Second, it may have been that there actually was a treatment effect, but there not adequate statistical power for the model to detect this effect. The power estimation techniques described above can be utilized to compute the specific effect sizes and available power for this data set. Using these methods, the effect size for the interaction term was approximately .38. This can be considered a large effect. Given an effect size of .38, a control group sample size $N = 111$, and a treatment group sample size $N = 75$ (e.g., the actual conditions of the analyses), there was a power of .90 to detect the presence of the interaction. This is a very high level of power to detect the effect of interest. Thus, if a treatment effect had not been identified, but the above calculations estimated a power of .90 to detect such an effect, then there is a low likelihood that a treatment effect was mistakenly not identified when it truly did exist. Note, though, that this level of power does not necessarily suggest that fewer numbers of subjects could be used in the future. There are additional complicating factors such as the stability of model estimation [e.g., the minimum necessary ratio of subjects to parameters (see Bentler & Chou, 1988)] that must also be considered when designing a study.

These calculations can also provide useful information for the planning of future studies. Say, for example, this study was to be replicated using a balanced design with $N = 75$ in the treatment group and $N = 75$ in the control group (thus taking away $N = 36$ control cases from the original analyses), the resulting power would be .85. Thus, although adding cases only to the control group does indeed raise the power, there is a curve of diminishing returns with regard to how many subjects are added. The addition of $N = 36$ subjects to the control group (from $N = 75$) increases the power by .05. Consider instead if $N = 36$ cases were added to the

treatment group so that the original sample sizes were reversed: $N = 111$ in the treatment group and $N = 75$ in the control group. The resulting power to detect the interaction effect rises only slightly to .91. This is an increase in power of .06 over the original sampling design. Finally, if a balanced design of $N = 111$ in the treatment group and $N = 111$ in the control group were to be used, there would be a power of .95 to detect the interaction effect.¹⁴ There are thus large differences in power depending upon whether cases are added just to the control group, just to the treatment group, or equally to both groups. These differences in power should be carefully considered when estimating sample size requirements for future studies.

DISCUSSION

There were two primary goals for the present paper. The first was to describe LC analysis, to identify the advantages of utilizing LC models in the evaluation of prevention and intervention programs, and to apply these models to both artificial and real data to demonstrate model building strategies and subsequent interpretations. The second goal was to discuss a new method for computing treatment effect sizes and for calculating specific power estimates to be used in the planning and design of future prevention and intervention studies.

LC Modeling in Intervention Research

Developmental theories often describe change in terms of continuous developmental growth trajectories over time (see, e.g., Cairns & Cairns, 1994; Caspi, Elder, & Bem, 1987; Cicchetti & Richters, 1993; Conduct Problems Prevention Research Group, 1992; Costello & Angold, 1993; Loeber *et al.*, 1993). For example, many development processes are described as having some point of onset followed by a rate of growth that is initially slow but then accelerates over some period of time, after which there is a period of deceleration, and finally, some plateau is achieved and subsequently maintained. Within this overall developmental trajectory there are many opportunities for the existence of individual differences: children differ in age at onset, in rates of acceleration and deceleration, and in levels of final plateau. It is often these individual differences in growth that are of greatest interest when studying development over time. For the intervention scientist, the questions of interest are often even

¹⁴Note that these differences in power as a function of sample size would be more pronounced if considering a level of initial power that was substantially smaller than .90.

more complex. It is first necessary to understand the characteristics of the normative developmental process as it exists naturally over time [(Cicchetti & Toth, 1992); or “baseline” modeling (Kellam *et al.*, 1991)]. Once some understanding of the natural development process is achieved, it is then necessary to evaluate the degree to which a prevention or intervention program served to alter or modify this natural developmental growth process over time. For example, if the normative age at onset for alcohol use is 12 years, can a prevention program be implemented to delay this initial starting point to 14 years of age? If the normative developmental trajectory of reading ability is one-fourth of a standard deviation per year for a disadvantaged child, can an intervention program accelerate this growth trajectory to one-half a standard deviation per year?

Although these are the types of questions that are of greatest interest when studying development and the malleability of development over time, traditional types of statistical analyses do not typically allow for the testing of these questions. Instead, more traditional regression-based analyses assess average individual standing relative to the group mean. For example, if a positive regression parameter is estimated between a Time 1 and a Time 2 measure of the same construct, this reflects that individuals above the mean at Time 1 tended to remain above the mean at Time 2 and individuals below the mean at Time 1 tended to remain below the mean at Time 2. No inference is made regarding whether the mean increased or decreased, but only that the average relative standing was similar at the two time points. This does not inform about growth or individual differences in growth (Rogosa, 1988; Rogosa & Willett, 1985).

It is important to note that there are instances in which fixed-effects models are wholly appropriate for the study of longitudinal stability and change over time (e.g., Diggle *et al.*, 1994; Dwyer, 1983). However, the conditions commonly encountered in intervention evaluation research, especially with children and adolescents, do not typically meet the assumptions of these fixed-effects models. Not only does this result in several statistically related problems, but fixed-effects models do not typically correspond to the basic tenets of developmental theory. For example, say that an intervention program is designed to accelerate the natural development of reading ability in disadvantaged children. The theoretically derived hypotheses are thus stated in terms of normative and altered developmental trajectories over time. However, the traditional types of analyses examine the data in terms of time-specific standing relative to the group mean. The theory is positing one question but the analytic technique is assessing a somewhat different question. This contradiction between question and answer can potentially limit the validity of the evaluation of the proposed hypotheses, which in turn limits the understanding of the theory in general.

LC analysis (and other forms of growth models as well) allow for the evaluation of proposed hypotheses in a way that is more consistent with the underlying theory of interest. The cornerstone of the LC model is the estimation of group-level parameters that define the characteristics of the continuous developmental trajectory over time and the estimation of individual variability around these group parameters. This individual variability can then be modeled as a function of treatment condition, initial status, demographic covariates, or other mediating or moderating influences of interest. The modeling strategies discussed here allow for the exploration of the normative developmental process that exists without exposure to the treatment. Once modeled, the growth in the treatment condition is compared to the natural developmental trajectory, and given certain methodological assumptions, differences between the two trajectories can be attributed to the intervention. Not only are the LC models more consistent with many of the underlying developmental theories of prevention and intervention science, but also these models have been shown to have significantly more power to detect an existing treatment effect. For these reasons, we strongly recommend that growth modeling techniques be closely considered when evaluating the efficacy of prevention and intervention programs.

Another advantage of LC models is that they may easily be modified to allow for the computation of treatment effect sizes and specific power estimates under a variety of assumptions and conditions. In this time of dwindling financial resources, it is becoming increasingly important to design successful intervention programs that are not only likely to identify a treatment effect if such an effect exists, but to do so in a cost-efficient manner. The techniques proposed here allow for a variety of design characteristics to be considered when calculating statistical power [see MacCallum, Browne, & Sugawara (1996) for a general alternative to the Satorra-Saris power estimation approach]. Once a given treatment effect size is defined, power can be computed as a function of the number cases in the control group, the number of cases in the treatment group, the number of total repeated measures obtained, and the number of time periods throughout which the repeated measures are applied. Knowledge of this range of information can help a researcher make an informed decision about the specific combination of design characteristics that will yield the optimal level of power at the lowest financial cost.

Limitations and Future Directions

Despite the advantages LC models provide to the study of development over time, there are several limitations of this technique as well.

First, as with all growth models, a minimum of three time points is required for proper estimation, and four or five time points are preferable. Further, unlike the HLM growth modeling framework, LC models currently require time-structured data in which all subjects are assessed within the same time points. The interval between assessments can vary across time (as in the Baltimore data examined earlier), but the intervals must be the same for all individuals. Also unlike the HLM modeling framework, current estimation techniques for LC models are rather limited in the analysis of missing data. New techniques are becoming increasingly available that allow for the modeling of missing and incomplete data that can be applied within the LC framework (Graham, Hofer, & MacKinnon, 1996; Muthén, Kaplan, & Hollis, 1987), but further work is needed in this area. Another limitation is that the maximum likelihood estimator used to evaluate these models makes the assumption of multivariate normality. Violations of this assumption have been shown to be problematic (Muthén & Kaplan, 1985, 1992), but recent advances in more robust methods of estimation have been shown to be quite promising (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1993), and several of these estimation methods could be applied within the LC framework. Finally, the LC techniques presented above do not consider potential clustering effects (e.g., children nested within classrooms, siblings nested within families). This is an important consideration when evaluating school-based intervention programs, and recent developments allow for the incorporation of clustered data within the LC model (McArdle & Hamagami, 1996; Muthén, 1994; *in press*).

The basic LC model can be extended in a number of exciting directions that were not discussed here. For example, instead of using composite manifest variables within each time point, a multiple-indicator latent factor can be defined to model the effects of measurement error. Second, as in the general structural equations model, mediators can be included to understand better the specific mechanisms associated with a particular treatment effect. These mediators can be modeled as time-specific influences on later developmental trajectories, or development can be modeled within the mediators themselves, and later growth in the outcome measure can be predicted from earlier growth in the mediator. Finally, all of the models presented above can be applied to quasi-experimental and observational data as well. For example, instead of defining the multiple groups as a function of treatment or control group membership, the groups could be defined based on the child's gender or on the child's parents' alcoholism diagnosis. Differences in the shape and variability of growth can thus be examined as function of naturally occurring grouping variables that do not involve random assignment to condition. See McArdle (1988, 1989, 1991)

and McArdle and Epstein (1987) for many additional examples of powerful and creative LC modeling applications.

In sum, we believe that there are a number of both statistical and theoretical advantages associated with the use of LC analysis in the evaluation of individual differences in rates of change over time, especially when evaluating developmental theories in intervention research. LC analysis provides yet another tool to the applied developmental researcher to be used in the quest to understand better stability and change in children over time.

REFERENCES

- Aiken, L. S., & West, G. W. (1991). *Multiple regression: testing and interpreting interactions*. Newbury Park, CA: Sage.
- Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419–456.
- Bock, R. D. (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Browne, M. W., & Du Toit, S. H. C. (1991). Models for learning data. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 47–68). Washington, DC: American Psychological Association.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data Analysis Methods*. Newbury Park, CA: Sage.
- Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education, 8*, 158–233.
- Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risks: Pathways of youth in our time*. New York: Cambridge University Press.
- Caspi, A., Elder, G. H., & Bem, D. J. (1987). Moving against the world: Life course patterns of explosive children. *Developmental Psychology, 23*, 308–313.
- Cicchetti, D., & Richters, J. (1993). Developmental considerations in the investigation of conduct disorders. *Development and Psychopathology, 5*, 331–344.
- Cicchetti, D., & Toth, S. L. (1992). The role of developmental theory in prevention and intervention. *Development and Psychopathology, 4*, 489–493.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conduct Problems Prevention Research Group (1992). A developmental and clinical model for the prevention of conduct disorder: The FAST Track program. *Development and Psychopathology, 4*, 509–527.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Costello, E. J., & Angold, A. (1993). Toward a developmental epidemiology of the disruptive behavior disorders. *Development and Psychopathology, 5*, 91–101.
- Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis*. The Stanford Evaluation Consortium, Occasional Papers, July.
- Cronbach, L. J., & Furby, L. (1970). How should we measure “change”—or should we? *Psychological Bulletin, 74*, 68–80.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- Curran, P. J., West, S. G., & Finch, J. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16–29.
- Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent alcohol use

- and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology*, 65, 130–140.
- Diggle, P. J., Liang, K. L., and Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford: Clarendon Press.
- Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford.
- Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31, 197–218.
- Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351–362.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.
- Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.
- Kellam, S. G., Werthamer-Larsson, L., Dolan, L. J., Brown, C. H., Mayer, L. S., Rebok, G. W., Anthony, J. C., Laudolff, J., Edelson, G., & Wheeler, L. (1991). Developmental epidemiologically based preventive trials: Baseline modeling of early target behaviors and depressive symptoms. *American Journal of Community Psychology*, 19, 563–584.
- Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, 35, 259–281.
- Laird, N. M., & Ware, H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.
- Loeber, R., Wung, P., Keenan, K., Giroux, B., Stouthamer-Loeber, M., Van Kammen, W. B., & Maughan, B. (1993). Developmental pathways in disruptive child behavior. *Development and Psychopathology*, 5, 103–133.
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- McArdle, J. J. (1986). Latent growth within behavior genetic models. *Behavioral Genetics*, 16, 163–200.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology*, 2nd ed. New York: Plenum Press.
- McArdle, J. J. (1989). Structural modeling experiments using multiple growth functions. In P. Ackerman, R. Kanfer, & R. Cudeck (Eds.), *Learning and individual differences: Abilities, motivation and methodology* (pp. 71–117). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McArdle, J. J. (1991). Structural models of developmental theory in psychology. In P. Van Geert & L. P. Mos (Eds.), *Annals of Theoretical Psychology, Vol. VII*. (pp. 139–160). New York: Plenum Press.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133.
- McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Meredith, W., & Tisak, J. (1984). "Tuckerizing" curves. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Muthén, B. O. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model*. Mooresville, IN: Scientific Software.
- Muthén, B. O. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: APA.
- Muthén, B. O. (1994). *Latent variable modeling of longitudinal and multilevel data*. Invited

- paper for the Showcase Session, Section on Methodology, American Sociological Association, Aug.
- Muthén, B. O. (in press.) Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In A. Sayer, & L. Collins (Eds.), *New methods for the analysis of change*. Washington, DC: APA.
- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2, 371–402.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–189.
- Muthén, B. O., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19–30.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52, 431–462.
- Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrika*, 51, 83–90.
- Rogosa, D. R. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics*, 12, 185–195.
- Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research*. New York: Springer.
- Rogosa, D., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, 50, 203–228.
- Saris, W. E., & Stronkhorst, H. (1984). *Causal modeling in nonexperimental research*. Amsterdam: Sociometric Research Foundation.
- SAS Institute, Inc. (1994). SAS/STAT user's guide. Cary, NC: SAS Institute, Inc.
- Satorra, A., & Saris, W. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 51, 83–90.
- Söbom, D. (1982). Structural equation models with structured means. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction*. Amsterdam: North-Holland.
- Stoolmiller, M., Duncan, T., Bank, L., & Patterson, G. (1993). Some problems and solutions in the study of change: Significant patterns in client resistance. *Journal of Consulting and Clinical Psychology*, 61, 920–928.
- Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23, 19–23.
- Willett, J. B. (1991). Measuring change: The difference score and beyond. In H. J. Walberg and G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation*. Oxford: Pergamon Press.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of change. *Psychological Bulletin*, 116, 363–381.
- Wohlwill, J. F. (1991). Relations between method and theory in developmental research: A partial-isomorphism view. *Annals of Theoretical Psychology*, 7, 91–138.