

General growth mixture modeling for randomized preventive interventions*

BENGT MUTHÉN†

Graduate School of Education & Information Studies, University of California, Moore Hall, Box 951521, Los Angeles, CA 90095-1521, USA
bmuthen@ucla.edu

C. HENDRICKS BROWN

University of South Florida, FL, USA

KATHERINE MASYN, BOOIL JO

University of California, Los Angeles, CA, USA

SIEK-TOON KHOO

Arizona State University, AZ, USA

CHIH-CHIEN YANG

National Taichung Teachers College, Taiwan

CHEN-PIN WANG

University of South Florida FL, USA

SHEPPARD G. KELLAM

Johns Hopkins University and American Institutes for Research, Baltimore, MD, USA

JOHN B. CARLIN

University of Melbourne, Australia

JASON LIAO

Medical University of South Carolina, SC, USA

SUMMARY

This paper proposes growth mixture modeling to assess intervention effects in longitudinal randomized trials. Growth mixture modeling represents unobserved heterogeneity among the subjects using a finite-mixture random effects model. The methodology allows one to examine the impact of an intervention on subgroups characterized by different types of growth trajectories. Such modeling is informative when examining effects on populations that contain individuals who have normative growth as well as non-normative growth. The analysis identifies subgroup membership and allows theory-based modeling of intervention effects in the different subgroups. An example is presented concerning a randomized

*This paper was presented at the 2000 Joint Statistical Meetings in Indianapolis, Indiana, August 13–17, 2000. It is a revised version of the 1997 paper *General Growth Mixture Modeling of Latent Trajectory Classes: Perspectives and Prospects*, an earlier version of which was presented at the Prevention Science Methodology Group meeting in Tempe, Arizona, May 22–24, 1996.

†To whom correspondence should be addressed

intervention in Baltimore public schools aimed at reducing aggressive classroom behavior, where only students who were initially more aggressive showed benefits from the intervention.

Keywords: Growth modeling; Latent variables; Maximum likelihood; Randomized trials; Trajectory classes; Treatment-baseline interaction.

1. INTRODUCTION

This paper presents a novel application of growth mixture modeling (Muthén and Shedden, 1999; Muthén, 2001a,b; Muthén and Muthén, 2000, 2001) to preventive intervention trials in which individuals are randomized into intervention and control groups and measured repeatedly before and after the start of the intervention. The strength of randomized repeated measures studies is that they allow the assessment of intervention effects on trajectories rather than merely focussing on overall intervention effects at a specific time point. The analysis better utilizes this strength by allowing for many forms of unobserved heterogeneity among subjects typically encountered in prevention studies, both with respect to development in the control group and with respect to the intervention effects. The analysis can also help point out advantageous refinements in the design of future intervention studies.

Development in the control group often needs to be described in terms of unobserved trajectory classes of development, within which there may be further individual trajectory variation. For example, some children in early school grades may be on a developmental path of reading disability, others may show mild forms of reading problems, while still others progress more normally (Muthén *et al.*, 2000). Another example involves different trajectories of aggressive/disruptive behavior. Evidence for the existence of three patterns of aggression trajectories—an early onset, a late onset, and a stable low aggressive pattern—has been reported by Moffitt (1993). A third example involves three major trajectory classes of alcohol drinking behavior among young adults with a normative low use class, an early onset class, and an escalating class (Muthén and Muthén, 2000). Multiple trajectories are often useful in medicine; Pearson *et al.* (1994) considered different groups of males with linear or exponential growth in prostate specific antigen (PSA). The average trajectories for the classes in these examples are different from one another with individual variation around each. It is important to be able to distinguish between individuals in the different classes because membership in different classes may have different antecedents, e.g. poverty for reading development, as well as consequences, e.g. alcohol dependence for more severe drinking behavior (Muthén and Shedden, 1999) and prostate cancer for those with exponential growth in PSA (Pearson *et al.*, 1994). This paper will study an example from randomized preventive field trials conducted in Baltimore by Johns Hopkins University, the Baltimore City Public Schools, and Morgan State University (Dolan *et al.*, 1993; Ialongo *et al.*, 1999). These studies intervene during first and second grade to improve reading and reduce aggression with outcomes assessed through middle school and beyond.

Section 2 gives a description of the Baltimore intervention study. Section 3 proposes two kinds of growth mixture models that allow for differential intervention impact among unobserved subgroups of subjects. Section 4 puts the models in a general framework and presents maximum-likelihood estimation using the EM algorithm. Section 5 shows the analysis results. Section 6 concludes.

2. THE BALTIMORE INTERVENTION STUDY

The motivation for the analyses is a school-based preventive intervention study carried out by the Baltimore Prevention Research Center under a partnership between the Johns Hopkins University, the Baltimore City Public Schools, and Morgan State University. In this intervention trial, children were followed from first to seventh grade with respect to the course of aggressive behavior, and a follow-up to

age 18 also allowed for the assessment of intervention impact on the probability of juvenile delinquency as indicated by juvenile court records.

One of the interventions applied during the first and second grade was the Good Behavior Game (GBG), a universal intervention aimed at reducing aggressive behavior. GBG is a classroom-based behavior management strategy for teachers that showed positive effects on short-term aggressive (Dolan *et al.*, 1993) and off-task behavior (Brown, 1993), as well as aggressive behavior in the long-term, i.e. through grade 7 (Kellam *et al.*, 1994). Key scientific questions address whether the GBG reduces the slope of the aggression trajectory across time, whether the intervention varies in impact for children who initially display higher levels of aggression, and whether the intervention impacts distal outcomes. It has been suggested that GBG may have its largest effect for those who are in the middle trajectory class, showing milder forms of problems, while not being strong enough to affect the most seriously aggressive children and not needed for members of the stable non-aggressive group. Analyses of these hypotheses are presented in this paper. Allowing for multiple trajectory classes in the growth model gives a flexible way to assess differential effects of the intervention. Intervention effects may differ across trajectory classes with respect to the rate of change over time and may also produce changes in trajectory class membership.

The overall design of the study involved random assignment of both schools and classrooms after making sure all first-grade classes within a school were balanced on kindergarten performance. Schools were first matched into six triplets and then randomly assigned within blocks to receive only the standard setting in all first-grade classrooms, to receive the GBG in one or more of its classes, or to receive a separate learning intervention in one or more of its classes. Within those schools where the GBG was made available, first-grade classrooms were randomly assigned to receive either this new intervention or the standard control setting condition. Further details on the design can be found in Brown and Liao (1999). For the purposes of this study, the analyses have been limited to the children receiving the GBG and their corresponding controls within the same schools. The primary outcome variable of interest was teacher ratings of each child's aggressive behavior in the classroom for grades 1–7. After an initial assessment in fall of first grade, the intervention was administered during the first two grades, with nearly all children remaining in the same intervention condition in the second year as they were in the first. Teacher ratings of a child's aggressive behavior were made from fall and spring for the first two grades and every spring in grades 3–7. The ratings were made using the Teacher's Observation of Classroom Adaptation Revised (TOCA-R) instrument (Werthamer-Larsson *et al.*, 1991), consisting of an average of 10 items, each rated on a six-point scale from 'almost never' to 'almost always'. Information was also collected on other concurrent and distal outcomes, including school removal and juvenile court records. The current analyses focus on boys and intervention status as defined by classroom assignment in fall of first grade, resulting in a sample of 119 boys in the intervention group and 80 boys in the control group.

3. GROWTH MIXTURE MODELING

To investigate whether or not subgroups of children benefit differently from the intervention, a finite-mixture random effects model will be formulated, where the unobserved subgroups of the mixture are conceptualized as different trajectory classes captured by a latent class variable with K classes. Two general growth mixture models will be studied.

3.1 Growth mixture model 1

Model 1 assumes that intervention effects are captured in the average slopes for each class. The notion is that an individual has a certain trajectory class membership that does not change over time. The intervention produces a change in within-class trajectory from that expected for controls.

Assume for individual i in class k ($k = 1, 2, \dots, K$),

$$y_{it} = \eta_{0i} + \eta_{1i} a_t + \eta_{2i} a_t^2 + \epsilon_{it}, \quad (1)$$

where y_{it} ($i = 1, 2, \dots, n$; $t = 1, 2, \dots, T$) are aggression outcomes influenced by the random effects η_{0i} , η_{1i} , and η_{2i} described below. The residuals ϵ_{it} have a $T \times T$ covariance matrix Θ_k , possibly varying across the trajectory classes ($k = 1, 2, \dots, K$). The intervention begins after the first measurement occasion. Setting $a_1 = 0$ in (1) defines η_{0i} as pre-intervention initial status at $t = 1$, i.e. fall of first grade. The remaining a_t values are set according to the distance in timing of measurements. It is assumed for simplicity in (1) that the a_t values do not vary across class or across intervention groups so that the growth function is the same.

Let the dummy variable I_i denote the intervention status for individual i ($I = 0$ for the control group and $I = 1$ for the intervention group). The random effects are allowed to have different distributions for individuals belonging to different trajectory classes and for different intervention status. For class k ,

$$\eta_{0i} = \alpha_{0k} + \zeta_{0i}, \quad (2)$$

$$\eta_{1i} = \alpha_{1k} + \gamma_{1k} I_i + \zeta_{1i}, \quad (3)$$

$$\eta_{2i} = \alpha_{2k} + \gamma_{2k} I_i + \zeta_{2i}. \quad (4)$$

The residuals ζ_i have a 3×3 covariance matrix Ψ_k , possibly varying across classes k ($k = 1, 2, \dots, K$). For simplicity, Ψ_k and Θ_k are assumed to not vary across intervention groups. As seen in (2)–(4), the control group ($I_i = 0$) consists of children from different trajectory classes that vary in the means of the growth factors, α_{0k} , α_{1k} , and α_{2k} . This represents the normative development in the absence of intervention. Because of randomization, the control and intervention group are assumed to be statistically equivalent at t_1 . This implies that I is assumed to have no effect on η_{0i} in (2) so that α_0 represents the mean of the initial-status random effect, common to both the control and intervention group. Intervention effects are described by γ_{1k} , γ_{2k} as a change in average growth rate that can be different for different classes k .

It may be noted that this model assumes that intervention status does not influence class membership. Alternative models were also pursued, however. Regressing class membership on intervention status, it was found that class sizes did not vary significantly across intervention groups. A technical report available from the first author includes a model that also allows transitions between classes as a function of the intervention.

3.2 Growth mixture model 2

Model 2 is the same as model 1, but adds a distal outcome that is influenced by the growth process for y . Consider, for example, a categorical outcome u . Model 2 assumes that the u probabilities are affected by the trajectory classes and that the intervention has a different effect on u for different trajectory classes.

With a binary distal outcome the class influence is described as the logit regression

$$\text{logit } P(u_i = 1 \mid \text{class } k, I_i) = -\tau_k + \kappa_k I_i. \quad (5)$$

Noting that $-\tau_k + \kappa_k I_i$ is the log odds for $u_i = 1$ versus $u_i = 0$ for individual i in class k , the intervention effect is expressed by the corresponding log odds ratio for $I_i = 1$ versus $I_i = 0$ in class k , obtained as the difference

$$-\tau_k + \kappa_k - (-\tau_k) = \kappa_k. \quad (6)$$

An odds ratio estimate and corresponding confidence interval are obtained by exponentiating the κ_k estimate and confidence limits.

The effect of class membership on the distal outcome can be expressed by the log odds for $u_i = 1$ versus $u_i = 0$ for individual i in class k , or by the corresponding log odds ratio for class k compared to a normative class K ,

$$-\tau_k + \kappa_k I_i - (-\tau_K + \kappa_K I_i). \quad (7)$$

It follows from (7) that when the intervention effect on the distal outcome is constant across classes, i.e. $\kappa_1 = \kappa_2, \dots = \kappa_K$, the log odds ratio for the distal outcome when comparing class k to class K is $-\tau_k + \tau_K$.

4. GROWTH MIXTURE MODELING FRAMEWORK, ESTIMATION, AND MODEL ASSESSMENT

The two growth mixture models proposed for the Baltimore intervention study may be seen as special cases of a more general modeling framework presented by Muthén and Shedden (1999) and extended by Muthén and Muthén (2001, see Appendix 8). Following is a brief review of this work as it pertains to the current models.

4.1 Modeling framework

The observed variables are \mathbf{x} , \mathbf{y} , and \mathbf{u} , where \mathbf{x} denotes a $q \times 1$ vector of covariates, \mathbf{y} denotes a $p \times 1$ vector of continuous outcome variables, and \mathbf{u} denotes an $r \times 1$ vector of binary outcome variables. In this application $r = 1$. The latent variable $\boldsymbol{\eta}$ denotes an $m \times 1$ vector of continuous variables and \mathbf{c} denotes a latent categorical variable with K classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{iK})'$, where $c_{ik} = 1$ if individual i belongs to class k and zero otherwise.

The latent classes of \mathbf{c} influence both \mathbf{y} and \mathbf{u} . Consider first the \mathbf{y} part of the model. Conditional on class k ,

$$\mathbf{y}_i = \boldsymbol{\Lambda}_k \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad (8)$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \mathbf{x}_i + \boldsymbol{\zeta}_i, \quad (9)$$

where the residual vector $\boldsymbol{\epsilon}_i$ is $N(\mathbf{0}, \boldsymbol{\Theta}_k)$ and the residual vector $\boldsymbol{\zeta}_i$ is $N(\mathbf{0}, \boldsymbol{\Psi}_k)$, both assumed to be uncorrelated with other variables. Conditional on class k , (8) and (9) form a conventional latent variable model (see, for example, Bollen (1989)), where the density $[\mathbf{y}_i | \mathbf{c}_i, \mathbf{x}_i]$ is $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where for class k ,

$$\boldsymbol{\mu}_i = \boldsymbol{\Lambda}_k (\boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \mathbf{x}_i), \quad (10)$$

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_k \boldsymbol{\Psi}_k \boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k. \quad (11)$$

A logistic regression is specified for the binary u . For class k ,

$$P(u_i = 1 | \mathbf{x}_i) = \frac{1}{1 + e^{\tau_k - \kappa_k' \mathbf{x}_i}}. \quad (12)$$

Translating model 1 and model 2 into matrix terms corresponding to the general model form, $\mathbf{x}_i = I_i$, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$, $\boldsymbol{\eta}_i = (\eta_{0i}, \eta_{1i}, \eta_{2i})'$, and

$$\Lambda_k = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & a_3 & a_3^2 \\ \vdots & \vdots & \vdots \\ 1 & a_T & a_T^2 \end{pmatrix}, \alpha_k = \begin{pmatrix} \alpha_{0k} \\ \alpha_{1k} \\ \alpha_{2k} \end{pmatrix}, \Gamma_k = \begin{pmatrix} 0 \\ \gamma_{1k} \\ \gamma_{2k} \end{pmatrix}. \quad (13)$$

With the modeling framework presented above, it is possible to examine a wide variety of hypotheses involving both the parameters and the dimensionality of \mathbf{c} and $\boldsymbol{\eta}$. This framework is an extension of the mixture mixed-effects model of Verbeke and Lesaffre (1996). It is also more general than the model of Nagin (1999); Roeder *et al.* (1999), and Jones *et al.* (2001): in their work $\boldsymbol{\Psi}_k = \mathbf{0}$, $\boldsymbol{\Theta}_k = \theta \mathbf{I}$. Identification of latent variable mixture models of the type presented here is demonstrated in Lubke *et al.* (2001). The modeling framework given above draws on that of Muthén and Shedden (1999) and Muthén and Muthén (2001, Appendix 8), which offers more generality than is needed here, including a multinomial regression of latent class membership on covariates, regressions among the random effects, time-varying covariates, multiple ordinal u variables following a latent class model, and partially known class membership. Applications to non-intervention settings are given in Muthén (2001a,b); Muthén and Muthén (2000), and Muthén *et al.* (2000). Applications to latent class membership representing nonparticipation (noncompliance) in intervention studies (Angrist *et al.*, 1996) are given in Jo (2000), Jo and Muthén (2000).

4.2 Estimation

With a sample of n independent observations on \mathbf{y} , \mathbf{u} , \mathbf{x} , the latent variable data $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_n$ and $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ may be viewed as missing data with the complete-data log likelihood conditional on \mathbf{x} expressed as

$$\sum_{i=1}^n (\log[\mathbf{u}_i | \mathbf{c}_i, \mathbf{x}_i] + \log[\boldsymbol{\eta}_i | \mathbf{c}_i, \mathbf{x}_i] + \log[\mathbf{y}_i | \mathbf{c}_i, \boldsymbol{\eta}_i, \mathbf{x}_i]), \quad (14)$$

where the first term is defined by (12), and the last two terms are normal densities. In this way, the bracket notation is used to refer to either probabilities or densities for simplicity in the presentation. Alternatively, with only $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ viewed as missing data, the complete-data log likelihood is

$$\sum_{i=1}^n (\log[\mathbf{u}_i | \mathbf{c}_i, \mathbf{x}_i] + \log[\mathbf{y}_i | \mathbf{c}_i, \mathbf{x}_i]), \quad (15)$$

The model can be estimated by maximum-likelihood using EM algorithms. Muthén and Shedden (1999) proposed an EM algorithm drawing on (14), while Muthén and Muthén (2001) use an EM algorithm drawing on (15). A brief summary of the latter approach follows.

Consider the conditional probability of individual i belonging to class k , given the observed data,

$$p_{ik} = P(c_{ik} = 1 | \mathbf{y}_i, \mathbf{u}_i, \mathbf{x}_i) = P(c_{ik} = 1) [\mathbf{y}_i | \mathbf{c}_i, \mathbf{x}_i] [\mathbf{u}_i | c_{ik} = 1, \mathbf{x}_i] / [\mathbf{y}_i, \mathbf{u}_i | \mathbf{x}_i]. \quad (16)$$

It follows that in (15),

$$\sum_{i=1}^n \log[\mathbf{y}_i | \mathbf{c}_i, \mathbf{x}_i] = \sum_{i=1}^n \sum_{k=1}^K c_{ik} \log[\mathbf{y}_i | \mathbf{x}_i]_k. \quad (17)$$

The EM algorithm used in Muthén and Muthén (2001) computes the expected value of \mathbf{c}_i using (16). Given this, the M step maximizes the expected complete-data log likelihood function, conditional on the observed data, separately for the \mathbf{y} , \mathbf{x} part of the model and the \mathbf{u} , \mathbf{x} part of the model. For the \mathbf{y} , \mathbf{x} part this is

$$E\left(\sum_{i=1}^n \log[y_i | \mathbf{c}_i, \mathbf{x}_i] | \mathbf{u}_i, \mathbf{y}_i, \mathbf{x}_i\right) = \sum_{i=1}^n \sum_{k=1}^K p_{ik} \log[y_i | \mathbf{x}_i]_k, \quad (18)$$

which corresponds to simultaneous estimation of the K groups with posterior-probability weighted sample mean vectors and covariance matrices. The maximization for the \mathbf{u} , \mathbf{x} part of the model is broken down into a multinomial regression optimization for \mathbf{c} related to \mathbf{x} (when this part of the model is present),

$$\sum_{n=1}^n \sum_{k=1}^K p_{ik} \log P(c_{ik} = 1 | \mathbf{x}_i) \quad (19)$$

and a logistic regression optimization for \mathbf{u} related to \mathbf{c} and \mathbf{x} ,

$$\sum_{i=1}^n \sum_{j=1}^r \sum_{k=1}^K p_{ik} \log P(u_{ij} = 1 | \mathbf{c}_i, \mathbf{x}_i). \quad (20)$$

This EM algorithm is implemented in the Mplus program (Muthén and Muthén, 2001), which is the program used for the analyses.[†] Mplus allows \mathbf{y} and \mathbf{u} to be missing at random (Little and Rubin, 1987). It should be noted that mixture models in general are prone to have multiple local maxima of the likelihood and the use of several different sets of starting values in the iterative procedure is strongly recommended.

4.3 Model assessment

For comparison of fit of models that have the same number of classes and are nested, the usual likelihood-ratio chi-square difference test can be used. Comparison of models with different numbers of classes, however, is accomplished by a Bayesian information criterion (BIC) (Schwartz, 1978; Kass and Raftery, 1993),

$$BIC = -2 \log L + r \ln n, \quad (21)$$

where r is the number of free parameters in the model. The lower the BIC value, the better the model.

The degree to which the latent classes are clearly distinguishable by the data and the model can be assessed by using the estimated conditional class probabilities for each individual. By classifying each individual into his/her most likely class, a $K \times K$ table can be constructed with rows corresponding to individuals classified into a given class. For individuals in each row, the column entries give the average conditional probabilities. This will be referred to as a classification table (Nagin, 1999). High diagonal and low off-diagonal values indicate good classification quality. A summary measure of the classification is given by the entropy measure (see, for example, Ramaswamy *et al.* (1993),

$$E_K = 1 - \frac{\sum_i \sum_k (-\hat{p}_{ik} \ln \hat{p}_{ik})}{n \ln K}, \quad (22)$$

where \hat{p}_{ik} denotes the estimated conditional probability for individual i in class k . Entropy values range from zero to one, where entropy values close to one indicate clear classifications in that the entropy decreases for probability values that are not close to zero or one.

[†]Input specifications for the Mplus analyses can be found at www.statmodel.com

The fit of the model to the data can be studied by comparing for each class estimated moments with moments created by weighting the individual data by the estimated conditional probabilities (Roeder *et al.*, 1999). To check how closely the estimated average curve within each class matches the data, it is also useful to randomly assign individuals to classes based on individual estimated conditional class probabilities. Plots of the observed individual trajectories together with the model-estimated average trajectory can be used to check assumptions (Bandein-Roche *et al.*, 1997).

5. GROWTH MIXTURE ANALYSES

In this section the Baltimore intervention data are analyzed in four steps: using a conventional single-class model; using an initial growth mixture exploration of the control and intervention groups; using model 1; and using model 2. Because children are clustered within classrooms, standard errors of parameter estimates were also estimated using a sandwich estimator assuming independent observations only across classrooms. The resulting standard errors were very similar to the unadjusted standard errors which are reported here.

5.1 *Conventional single-class analyses*

As a first step in the repeated measures analysis it is useful to study the normative development of aggressive behavior shown in the control group. This establishes the trajectory shape in the absence of intervention so that effects of the intervention can be more clearly understood. Initial explorations pointed to a quadratic growth curve model. The random effects did not need to be correlated. The time-specific residuals needed to be correlated for fall and spring for each of the two first grades. Likelihood-ratio chi-square testing was used to aid these decisions.

A joint analysis of the 80 control group children and the 119 intervention group children using a single-class ($K = 1$) version of the model of (1)–(4), i.e. a conventional Laird and Ware (1982) model, resulted in an insignificant intervention effect with the estimates (s.e.) $\hat{\gamma}_1 = -0.01$ (0.08), $\hat{\gamma}_2 = 0.00$ (0.01). This produces a spring grade 7 estimated mean difference between the control and intervention group of only -0.04 , or approximately 0.03 of the aggression score standard deviation at that time point, an inconsequential effect size.

5.2 *Initial growth mixture analyses*

An initial exploration by growth mixture analysis is important because model 1 includes many possible alternatives. The control group is first analyzed separately to establish normative growth in the absence of an intervention, followed by a separate analysis of the intervention group. Alternative variance assumptions were investigated, holding all variances equal across classes versus letting the intercept and residual variances differ across certain classes. Based on likelihood-ratio chi-square testing in the control group as well as the intervention group, it was found that the intercept and residual variances needed to be different for a class of children with stable low level of aggression. As a guide in choosing between models with different number of classes, the BIC was used. It is useful to determine the number of classes in separate analyses of the two groups for two reasons. First, the control group analysis suggests the number of classes in the absence of an intervention and given that it is assumed that the intervention does not influence class membership, this number of classes should also hold in the intervention group. Second, the joint analysis of the two groups adds its own model specifications and it is valuable to establish the number of classes without adding these specifications.

The BIC values were obtained for 1–5 classes for full and partial variance homogeneity in the control group and indicated a superior fit when allowing non-invariant variances. The BIC values suggested a

considerably better fit when allowing more than one class. With heterogeneous variances, the lowest value was at 4 classes although the 3-, 4-, and 5-class solutions had rather similar values.

The left column of Figure 1 shows the estimated mean growth curves for the 3-, 4-, and 5-class models for the control group.

The 3-class solution has class probabilities 0.09, 0.52, 0.39, the 4-class solution has class probabilities 0.08, 0.38, 0.41, 0.13, and the 5-class solution has class probabilities 0.08, 0.32, 0.45, 0.09, 0.06. Going from three to five classes provides an increasingly elaborate description of the trajectories, while the previously obtained classes do not change much when adding a new class. The 3- and 4-class solutions will be highlighted here. The two solutions share three of the classes and they will be named High, Medium, and Low corresponding to their relative positions. The remaining class in the 4-class solution will be named Late-starters.

Considering the 4-class solution, the 8% in the High class show a high aggression level in early grades that decreases over time. In line with Moffitt (1993) this group corresponds to an 'early starter' group of aggressive boys. The Late-starters class contains 13% of the children, showing a low initial aggression level that increases over time. The Medium class and the Low class have the highest probabilities, 38 and 41% respectively, and show low aggression trajectories that do not increase or decrease over time. The Low class has low intercept and residual variances indicating little fluctuation in the development. The Low class contains the stable low aggressive children.

The right column of Figure 1 shows the estimated mean curves for the intervention group using the 3-, 4-, and 5-class models. Here, the High class shows a decline earlier than for the control group, indicating a beneficial intervention effect on the highest-risk boys. A beneficial intervention effect is also indicated for the Late-starters class of the 4- and 5-class solutions. These intervention effects will now be examined in a joint analysis of control and intervention children.

5.3 Model 1 analyses

In Model 1, the joint analysis of the control and the intervention group based on the model in (1)–(4) uses the specifications arrived at from the initial analyses discussed above. Both the 3- and 4-class versions are studied for comparison.

In the joint analysis using four classes it was found that not only the Low class, but also the Late-starters class, required a separate specification of variance parameters. The Late-starters class was found to have insignificant intercept variance and significantly smaller residual variances than the High and Medium classes. The resulting 4-class solution had a log likelihood value of -1554.34 with the BIC value of 3394.53 (54 parameters). These can be compared to those of the 3-class model: -1604.40 and 3420.52 (40 parameters), respectively. The parameter estimates for the 4-class solution are shown in Table 1.

The estimated mean curves for the 3- and 4-class versions of Model 1 are shown in Figure 2. For the 3-class solution there appears to be a beneficial intervention effect for the High class through a lowered aggression trajectory. Although the difference in the means of the linear terms for the High class is sizable, $\hat{\gamma}_1 = -0.45$, this is not significantly different from zero (95% CI: $-0.92, 0.02$). The Medium class and the Low class show no intervention effects. Thus this model suggests that the intervention affects the group with the highest risk, but does not provide unequivocal evidence.

The intervention impact in the 4-class solution shows a pattern for the High class similar to that of the 3-class solution, but the result is again insignificant. A beneficial effect is also suggested for the Late-starters class, but is also non-significant. Overall, the likelihood-ratio test of any intervention effect in terms of the linear and quadratic means (γ_1, γ_2 coefficients) over the four classes is very small (a likelihood-ratio test gives $\chi^2(8) = 1.40$; $p > 0.50$ for the 4-class solution). The lack of significance is perhaps in large part due to low power given small class sizes in combination with large within-class

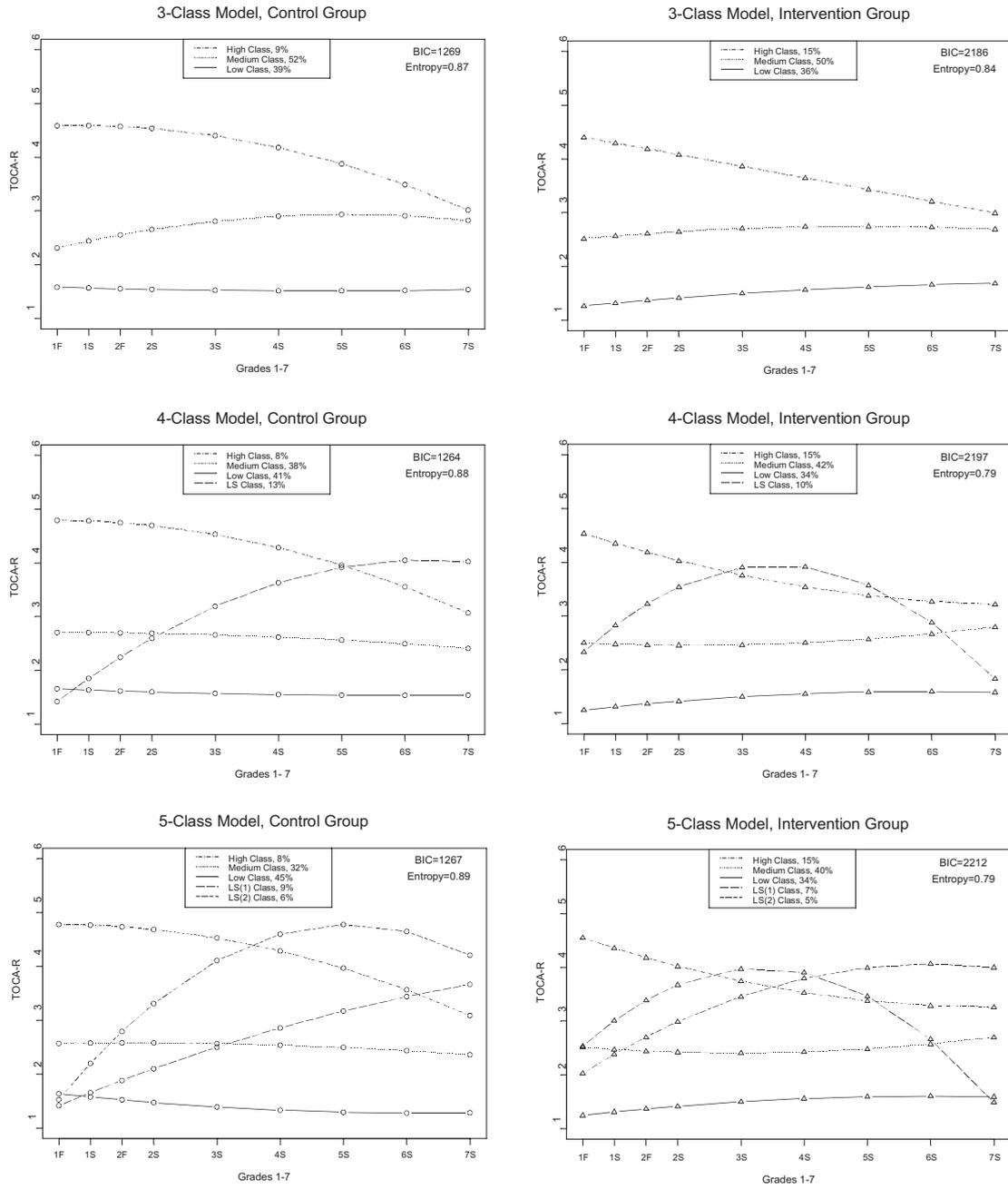


Fig. 1. Separately estimated mean growth curves for 3-, 4-, and 5-class models for control and intervention groups.

Table 1. Parameter estimates for 4-class model 1

Model	$y_{it} = \eta_{0i} + \eta_{1i}a_t + \eta_{2i}a_t^2 + \varepsilon_{it} \quad a_t = 0, 0.5, 1.0, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5$							
	$\eta_{0i} = \alpha_{0k} + \zeta_{0i}$							
	$\eta_{1i} = \alpha_{1k} + \gamma_{1k}I_i + \zeta_{1i}$							
	$\eta_{2i} = \alpha_{2k} + \gamma_{2k}I_i + \zeta_{2i}$							
	$V(\zeta \text{class } k) = \Psi_k$							
	$V(\varepsilon \text{class } k) = \Theta_k$							
	$P(c_{ik}) = \frac{e^{\alpha_{ck}}}{\sum e^{\alpha_{ck}}}$							
Aggression Growth Estimates								
Parameter	High class		Medium class		Low class		LS class	
	estimate (S.E.)		estimate (S.E.)		estimate (S.E.)		estimate (S.E.)	
α_0	3.846	(0.256)	2.571	(0.108)	1.531	(0.079)	1.382	(0.059)
α_1	0.502	(0.204)	0.076	(0.109)	-0.144	(0.049)	0.272	(0.071)
α_2	-0.078	(0.034)	-0.015	(0.018)	0.017	(0.049)	-0.014	(0.014)
γ_1	-0.329	(0.217)	-0.045	(0.117)	-0.079	(0.038)	-0.074	(0.089)
γ_2	0.025	(0.040)	0.015	(0.021)	0.015	(0.006)	-0.004	(0.017)
$V(\zeta_0)$	0.077	(0.042)	0.077	(0.042)	0.000	(fixed)	0.000	(fixed)
$V(\zeta_1)$	0.002	(0.001)	0.002	(0.001)	0.002	(0.001)	0.002	(0.001)
$V(\zeta_2)$	0.000	(fixed)	0.000	(fixed)	0.000	(fixed)	0.000	(fixed)
$V(\varepsilon_{1F})$	1.163	(0.176)	1.163	(0.176)	0.221	(0.057)	0.141	(0.029)
$V(\varepsilon_{1S})$	0.700	(0.129)	0.700	(0.129)	0.175	(0.037)	0.189	(0.040)
$V(\varepsilon_{2F})$	0.670	(0.111)	0.670	(0.111)	0.321	(0.078)	0.217	(0.044)
$V(\varepsilon_{2S})$	0.744	(0.119)	0.744	(0.119)	0.237	(0.053)	0.328	(0.073)
$V(\varepsilon_{3S})$	1.266	(0.243)	1.266	(0.243)	0.018	(0.007)	0.281	(0.089)
$V(\varepsilon_{4S})$	0.855	(0.146)	0.855	(0.146)	0.047	(0.018)	0.551	(0.149)
$V(\varepsilon_{5S})$	0.678	(0.129)	0.678	(0.129)	0.081	(0.029)	0.475	(0.139)
$V(\varepsilon_{6S})$	1.269	(0.213)	1.269	(0.213)	0.050	(0.021)	0.763	(0.214)
$V(\varepsilon_{7S})$	1.091	(0.200)	1.091	(0.200)	0.023	(0.014)	0.655	(0.191)
$C(\varepsilon_{1F}, \varepsilon_{1S})$	0.141	(0.030)	0.141	(0.030)	0.141	(0.030)	0.141	(0.030)
$C(\varepsilon_{2F}, \varepsilon_{2S})$	0.219	(0.048)	0.219	(0.048)	0.219	(0.048)	0.219	(0.048)

Latent class estimates

Parameter	Estimate	S.E.
α_{c1}	-0.395	0.302
α_{c2}	-0.190	0.287
α_{c3}	0.672	0.228
α_{c4}	0.000	(fixed)

variation; for example, the High class contains 15% of the sample, or only 12 boys from the control group and 18 boys from the intervention group (the within-class variation is shown in Figure 4).

Although the intervention effect is not significant, the estimated mean curves of Figure 2 show that for the High class, the estimated effect size is about one aggression score standard deviation for grades 2–6 in both the 3- and 4-class versions of the model. The High class contains about the same percentage of children in both the 3- and 4-class solutions, 14–15%. This is roughly comparable to the percentages of children found in the separate-group solutions of Figure 1.

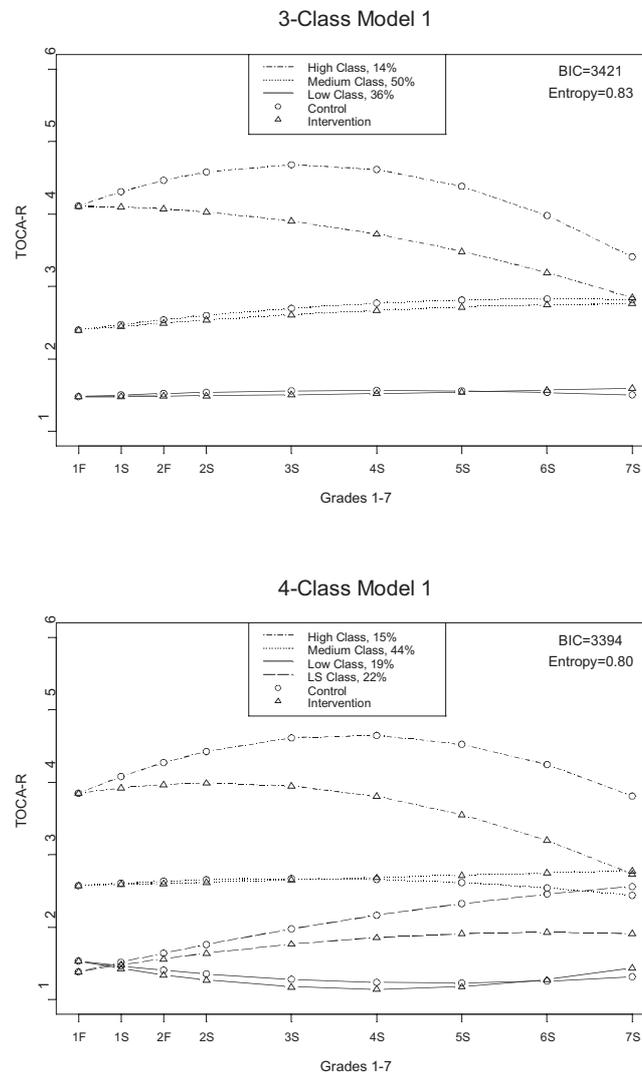


Fig. 2. Estimated mean growth curves for 3- and 4-class models.

In line with Section 4.3, the quality of the classification can be studied in terms of estimated probabilities in the classification table shown in Table 2, each row corresponding to individuals who are most likely to be in the particular class of that row. High classification quality is indicated by high diagonal probability values. Table 2 shows the results for the 4-class solution. The entropy value is 0.83 for the 3-class version and 0.80 for the 4-class version.

Figure 3 shows that the estimated 4-class model appears to fit the data well when compared to the probability-weighted means and variances. An exception is seen in the variances for the control group in grade 1.

A visualization of how the model matches the individual data is given in Figure 4 for the 4-class solution. As discussed in Section 4.3, this may be obtained by comparing the estimated mean curve in

Table 2. Classification table for 4-class model 1

Most likely class	Average posterior probabilities			
	High class	Medium class	Low class	LS class
High	0.864	0.136	0.000	0.000
Medium	0.057	0.903	0.017	0.024
Low	0.000	0.028	0.913	0.059
LS	0.001	0.064	0.092	0.844

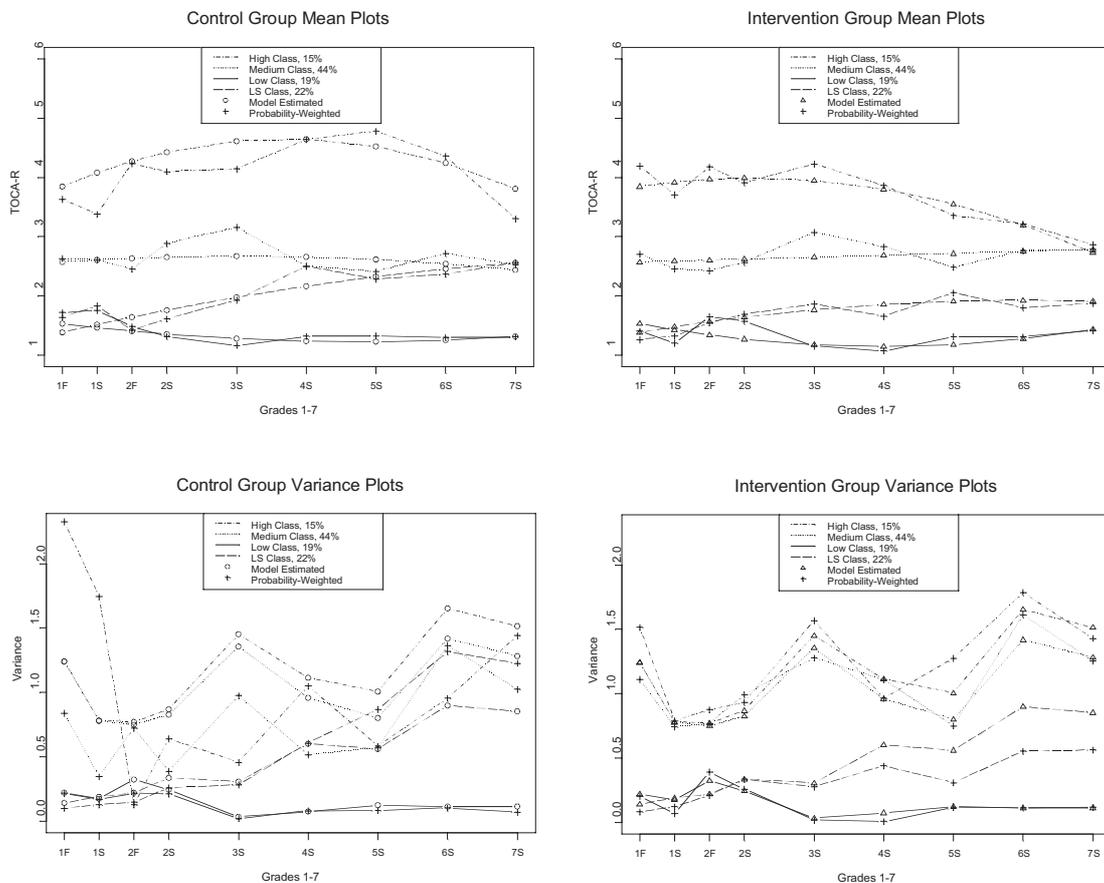


Fig. 3. Probability-weight means and variances for 4-class model.

each class to raw data trajectories for individuals assigned to that class by a random draw according to the estimated individual class probabilities. Figure 4 indicates that although individual trajectories fluctuate greatly, the estimated mean trajectories in the classes cut through the middle of the collection of individual trajectories rather well. Also, the smaller variances in the Low class and the Late-Starters class are evident in the figure.

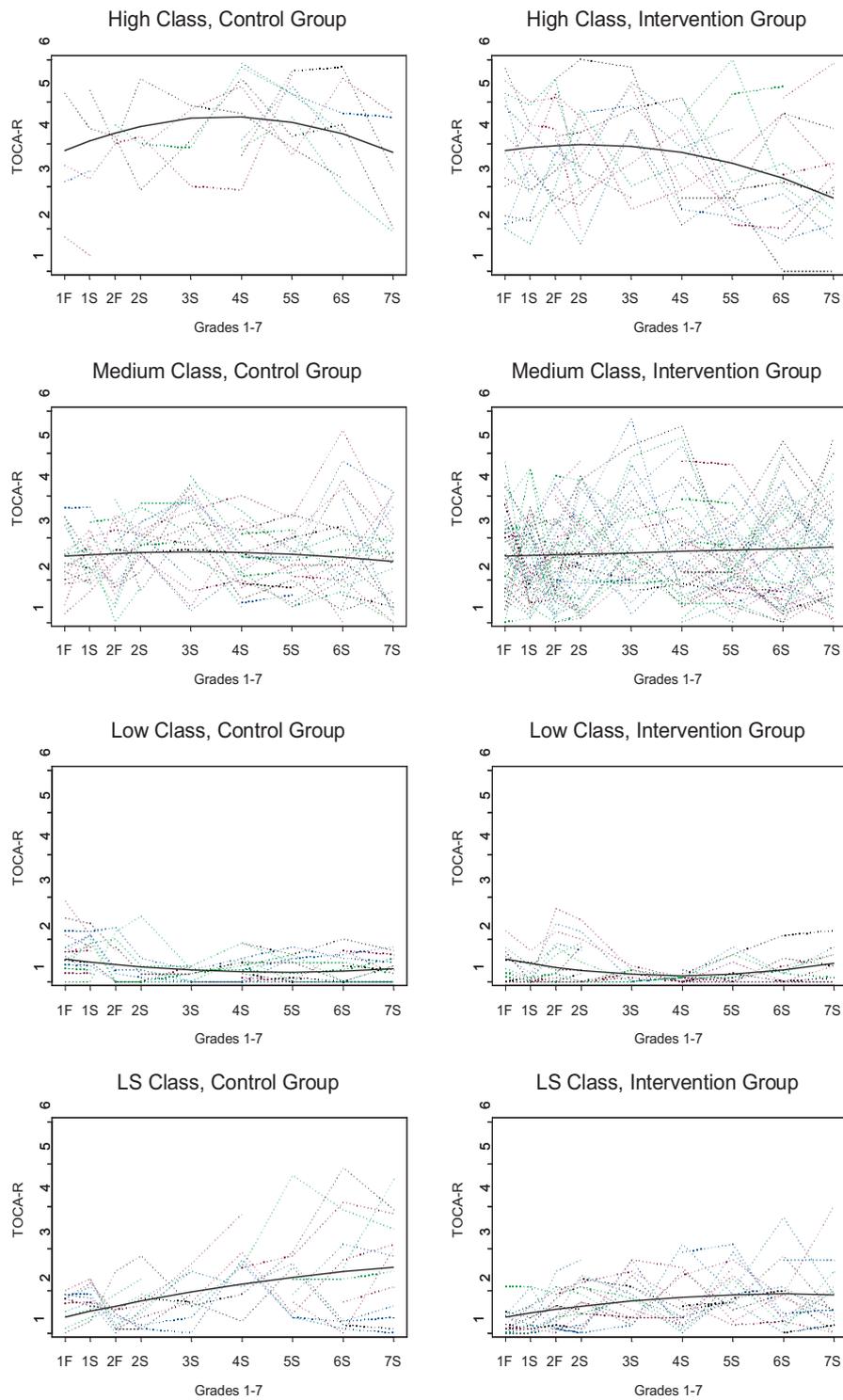


Fig. 4. Estimated mean growth curves and observed trajectories for 4-class model.

5.4 Model 2 analyses

Model 2 adds the distal binary outcome of juvenile delinquency prior to age 18. Only the 4-class version of the model is reported here. The estimated mean curves are essentially the same as for model 1. Two versions of model 2 were used, depending on whether the effect of intervention on the distal outcome of juvenile court record was allowed to vary across the classes (model 2a), or not (model 2b). The classification table is similar to that of model 1 with an entropy value of 0.81 for model 2a.

The estimated odds ratios based on the results from model 2a indicate positive intervention effects on juvenile delinquency in the High, Low, and Late-starter classes. However, none of the classes show a significant relationship between intervention status and the distal outcome at the 5% level. Comparing models 2a and 2b, the chi-squared difference is 1.96 with three degrees of freedom and a corresponding $p > 0.50$. Thus, class invariance for the effects of the intervention on juvenile delinquency cannot be rejected. Based on the model 2b results, the estimated common odds ratio for juvenile delinquency comparing the GBG group to the control group is 0.61 with a corresponding 95% confidence interval of (0.32, 1.14). While representing a positive intervention effect, the effect is not significant at the 5% level.

It is also possible to assess the effects of class membership on the distal outcome. Based on model 2b, boys in the High class are at a significantly higher risk for having a juvenile court record compared to boys in the Low class: estimated odds ratio is 8.11 (2.35, 27.97). Boys in the Late-starter and Medium class do not show a significantly increased risk.

6. CONCLUSIONS

This paper has discussed growth mixture modeling to assess intervention effects in randomized trials. The two model types that were proposed indicate some of the flexibility of the new methodology and serve as a stimulus for formulating other models. The methodology allows one to examine in detail the impact of an intervention on unobserved subgroups characterized by different types of growth trajectories. The analysis identifies subgroup membership and allows different intervention effects in the different subgroups. In addition, the analysis can predict the influence of subgroup membership on distal outcomes.

The growth mixture models described in this paper provide representative examples of how to determine worthwhile benefits from an intervention and when these effects are likely to appear. In this way, the growth mixture modeling becomes a powerful analytic tool when applied to randomized trials as well as to non-experimental research. The techniques illustrated here can be easily expanded to fit particular substantive hypotheses. For example, model 1 alternatives can examine the number of classes, the differential intervention effects on each class mean and variance, as well as basic assumptions such as balance in intervention and control at baseline. Variations of model 2 allow us to test differential effects across classes on distal outcomes as well as indirect effects of the intervention through mediators' latent classes. Further model variations are described in a technical report available from the first author.

As a caveat, it should be noted that these techniques should not be used as a substitute for reporting significant overall or population level effects. In fact, routine reliance on growth mixture modeling in the absence of main effects is likely to result in spurious findings because of the multiple comparisons problem. It is recommended that growth mixture modeling be carried out by comparing the empirical trajectories with those from existing empirical data or theory. In the current situation, the models produced results that explained previously published finding that pointed to short-term impact on multiple measures for those boys who began first grade with high levels of aggression (Dolan *et al.*, 1993) and significant benefit at sixth grade (Kellam *et al.*, 1994).

The idea of detecting different intervention effects for individuals belonging to different trajectory classes has important implications for designing future intervention studies. It is possible to select different interventions for individuals belonging to different trajectory classes using longitudinal screening

procedures. One may attempt to classify individuals into their most likely trajectory class based on a set of initial repeated measurements before the intervention starts. Alternatively, one may administer a universal intervention and follow up with a targeted intervention for individuals who show little or no intervention effect (Brown and Liao, 1999).

ACKNOWLEDGEMENT

The research of BM and his colleagues was supported under grant 1 K02 AA 00230-01 from NIAAA, by grant 1 R21 AA10948-01A1 from NIAAA, and by NIMH under grant No. MH40859. Work on this paper by CHB and his colleagues was supported by both NIMH and NIDA under grant MH40859, and by NIMH under grants No. MH01259, No. MH38725, and No. MH42968, and by NICHD under grant No. HD040051. The work has benefitted from continued discussion of the Prevention Science and Methodology Group, including presentations by Lawrence Mayer, and Muthén's Research Apprenticeship Course.

REFERENCES

- ANGRIST, J. D., IMBENS, G. W. AND RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.
- BANDEEN-ROCHE, K., MIGLIORETTI, D. L., ZEGER, S. L. AND RATHOUZ, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association* **92**, 1375–1386.
- BOLLEN, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.
- BROWN, C. H. (1993). Statistical methods for preventive trials in mental health. *Statistics in Medicine* **12**, 289–300.
- BROWN, C. H. AND LIAO, J. (1999). Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiologic perspective. *American Journal of Community Psychology*, special issue on prevention science **27**, 673–709.
- DOLAN, L., KELLAM, S. G., BROWN, C. H., WERTHAMER-LARSSON, L., REBOK, G. W., MAYER, L. S., LAUDOLFF, J., TURKKAN, J. S., FORD, C. AND WHEELER, L. (1993). The short-term impact of two classroom based preventive intervention trials on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology* **14**, 317–345.
- IALONGO, L. N., WERTHAMER, S., KELLAM, S. K., BROWN, C. H., WANG, S. AND LIN, Y. (1999). Proximal Impact of Two First-Grade Preventive Interventions on the Early Risk Behaviors for Later Substance Abuse, Depression and Antisocial Behavior. *American Journal of Community Psychology* **27** Vol. **5**, 599–641.
- JO, B. (2000). Estimation of intervention effects with noncompliance: alternative model specifications. *Forthcoming in Journal of Educational and Behavioral Statistics*.
- JO, B. AND MUTHÉN, B. (2000). Intervention studies with noncompliance: Complier Average Causal Effect Estimation in Growth Mixture Modeling. In Duan, N. and Reise, S. (eds), *Multilevel Modeling: Methodological Advances, Issues, and Applications*, Multivariate Applications Book Series, Lawrence Erlbaum Associates, To appear.
- JONES, B. L., NAGIN, D. S. AND ROEDER, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research* **29**, 374–393.
- KASS, R. E. AND RAFTERY, A. E. (1993). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.
- KELLAM, S. G., REBOK, G. W., IALONGO, N. AND 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**, 359–382.

- LAIRD, N. M. AND WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- LITTLE, R. J. AND RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- LUBKE, G., MUTHÉN, B. AND LARSEN, K. (2001). Global and local identifiability of factor mixture models. University of California, Los Angeles. Under review.
- MOFFITT, T. E. (1993). Adolescence-limited and life-course persistent antisocial behavior. *Psychological Review* **100**, 674–701.
- MUTHÉN, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In Collins, L. M. and Sayer, A. (eds), *New Methods for the Analysis of Change*, Washington, DC: APA, pp. 291–322.
- MUTHÉN, B. (2001b). Latent variable mixture modeling. In Marcoulides, G. A. and Schumacker, R. E. (eds), *New Developments and Techniques in Structural Equation Modeling*, Mahan, NJ: Lawrence Erlbaum Associates, pp. 1–33.
- MUTHÉN, B. AND MUTHÉN, L. (2000). Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*. **24**, 882–891.
- MUTHÉN, B. AND MUTHÉN, L. (2001). *Mplus User's Guide*. Los Angeles, CA: Muthén and Muthén.
- MUTHÉN, B. AND SHEDDEN, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55**, 463–469.
- MUTHÉN, B., KHOO, S. T., FRANCIS, D. AND KIM BOSCARDIN, C. (2000). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In Reise, S. R. and Duan, N. (eds), *Multilevel Modeling: Methodological Advances, Issues, and Applications*, (in press). Mahaw, NJ: Lawrence Erlbaum Associates.
- NAGIN, D. S. (1999). Analyzing developmental trajectories: a semi-parametric, group-based approach. *Psychological Methods* **4**, 139–157.
- PEARSON, J. D., MORRELL, C. H., LANDIS, P. K., CARTER, H. B. AND BRANT, L. J. (1994). Mixed-effect regression models for studying the natural history of prostate disease. *Statistics in Medicine* **13**, 587–601.
- RAMASWAMY, V., DESARBO, W., REIBSTEIN, D. AND ROBINSON, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science* **12**, 103–124.
- ROEDER, K., LYNCH, K. G. AND NAGIN, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association* **94**, 766–776.
- SCHWARTZ, (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.
- VERBEKE, G. AND LESAFFRE, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91**, 217–221.
- WERTHAMER-LARSSON, L., KELLAM, S. G. AND WHEELER, L. (1991). Effect of first-grade classroom environment on child shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology* **19**, 585–602.

[Received February 21, 2001; first revision August 12, 2001; second revision October 30, 2001;
accepted for publication October 30, 2001]