

Analyzing Criminal Trajectory Profiles: Bridging Multilevel and Group-based Approaches Using Growth Mixture Modeling

Frauke Kreuter · Bengt Muthén

Published online: 3 November 2007
© Springer Science+Business Media, LLC 2007

Abstract Over the last 25 years, a life-course perspective on criminal behavior has assumed increasing prominence in the literature. This theoretical development has been accompanied by changes in the statistical models used to analyze criminological data. There are two main statistical modeling techniques currently used to model longitudinal data. These are growth curve models and latent class growth models, also known as group-based trajectory models. Using the well known Cambridge data and the Philadelphia cohort study, this article compares the two “classical” models—conventional growth curve model and group-based trajectory models. In addition, two growth mixture models are introduced that bridge the gap between conventional growth models and group-based trajectory models. For the Cambridge data, the different mixture models yield quite consistent inferences regarding the nature of the underlying trajectories of convictions. For the Philadelphia cohort study, the statistical indicators give stronger guidance on relative model fit. The main goals of this article are to contribute to the discussion about different modeling techniques for analyzing data from a life-course perspective and to provide a concrete step-by-step illustration of such an analysis and model checking.

Keywords Latent class growth modeling · Growth mixture modeling · Zero-inflated Poisson distribution · Developmental trajectory groups

Introduction

Research on the relationship between age and criminal behavior continues to play a prominent role in the criminological literature (for overviews see Sampson and Laub 2005a; Piquero et al. 2003; Piquero 2007). The increasing availability of individual-level

F. Kreuter (✉)
Joint Program in Survey Methodology, University of Maryland, 1218 Lefrak Hall, College Park,
MD 20742, USA
e-mail: fkreuter@survey.umd.edu

B. Muthén
University of California, Los Angeles, CA, USA

longitudinal data in criminology (e.g., Elliot 1985; Tracy et al. 1990; Farrington and West 1993; Nieuwebeerta and Blokland 2003) has led to a larger number of empirical papers in recent years with individual “criminal careers” as the outcome of interest. Those articles address questions regarding patterns of criminal offending across the (full or partial) age range of the life course (e.g., Piquero and Buka 2002; Laub and Sampson 2003), the existence of typical offender trajectories (e.g., Nagin et al. 1995; D’Unger et al. 1998), possible correlates of or explanations for different offender trajectories (e.g., Blokland et al. 2005), and the prediction of future offending from juvenile offender trajectories in combination with certain covariates (Piquero and Buka 2002).

These studies, while different in content, all rely on analyses of the age-crime relationship. These respective analyses, as well as those that attempt to link other variables to the observed age-crime trajectories, require a decision on the statistical model used to estimate the trajectories of interest. Currently, two statistical modeling techniques are in prominent use for modeling longitudinal data in criminology: growth curve models (e.g., Raudenbush and Bryk 2002) and latent class growth models, also known as group-based trajectory models (e.g., Nagin and Land 1993; Roeder et al. 1999).

In a growth curve model, the joint distribution of the observed outcome variables (in this case, the number of convictions at each time point) is characterized as a function of age. Individual variation is expressed as random effects or growth factors that are allowed to vary across individuals, assuming a normal distribution. Like growth curve models, latent class growth analysis (LCGA) models, or group-based trajectory models,¹ model the development of criminal behavior as a function of age. However, instead of assuming normality for random effects, LCGA uses a small number of groups to approximate the distribution of developmental pathways across individuals. Group-based trajectory models are employed to study criminological theories that predict prototypical developmental etiologies and trajectories within the population (Nagin 1999, 140). These theories distinguish, for example, between adolescent-limited and persistent offenders (Moffitt 1993). Concretely, the use of group-based trajectory models specifies that the variation of the coefficients across individuals can be fully explained by group membership. The groups will differ in their developmental pathways, but, according to the model, there is no further variation within the group, i.e., all members of a group have the same expected outcome trajectory.

The increasing number of criminological articles that use latent class growth analysis have led some to voice concern about the implicit notion behind them, i.e., the existence of a finite number of distinct developmental trajectories and the classification of people into those trajectories. “A key question [...] is whether trajectory groups actually exist. If they do, then group-based trajectory modeling is clearly an ideal modeling approach. But if they do not, [...] we must ask what insight is gained and also what is lost in this kind of approximation” (Raudenbush 2005, p. 132). “Perhaps we are better off assuming continuously varied growth a priori and therefore never tempting our audience to believe [...] that groups of persons actually exist” (Raudenbush 2005, p. 136). But even if there is agreement on the existence of different groups (and with them typical offense trajectories), the question remains if those groups are distinct in their power to predict future criminal careers (Sampson and Laub 2005b, p. 907) and to which extent group members ‘follow’ the group trajectories (Sampson and Laub 2005b, p. 908).

¹ This approach is often referred to under its technical SAS procedure name PROC TRAJ (Jones et al. 2001).

We think that this discussion would be well served by reflecting on the various modeling alternatives. To this end, we compare not only growth curve modeling and group-based modeling, but expand that comparison to include growth mixture modeling (GMM) and non-parametric growth mixture modeling (NP-GMM). These additional modeling approaches are conceptually located between the two extremes usually discussed in the criminological literature, i.e., the assumptions of purely random versus purely group-based variation among individuals in their developmental trajectories. In this article, we will compare these four modeling approaches using a framework in which each approach is treated as a specific case of general growth mixture model.

The focus of this article is on the specification and interpretation of the different modeling alternatives as well as their relative performance for particular data sets rather than on their mathematical details or further substantive implications. To this end, we will illustrate the methods by analyses of data in common use in the criminological literature, the Cambridge data and the Philadelphia cohort study. These data are described in the next section and are almost entirely in the public domain.² For the Cambridge data the different mixture models yield quite consistent inferences regarding the nature of the underlying trajectories of convictions. For the Philadelphia cohort study the statistical indicators give stronger guidance on relative model fit, and point to GMM as the preferred model.

Population-based Conviction Data

We will use data from the “Cambridge Study” for the step-by-step demonstration of the different models, and show the modeling results for the data from the “Philadelphia Cohort Study”. Both data sets will also be used to further examine sensitivity in the model comparisons.

The so-called “Cambridge Study” includes data on 411 males from Cambridge, England who were followed from ages 10 to 40.³ This data collection effort, initiated by Dr. Donald J. West and continued by Dr. David Farrington, was undertaken to test several hypotheses about delinquency (Farrington and West 1990). The data set includes a rich set of covariates collected to determine the causes of crime and delinquency. The Philadelphia cohort study includes information on 13,160 males born in Philadelphia in 1958. Annual counts of police contacts are available from ages 4 to 26 for this birth cohort. A detailed discussion of these data can be found in various places (e.g., Tracy et al. 1990) and will not be repeated here.

Both data sets are typical in presenting two challenges that are characteristic for population-based conviction data where the prevalence of the non-normative behavior is, by definition, low. First, the outcome variable, *number of convictions* in the Cambridge study and *number of police contacts* in the Philadelphia study, is extremely skewed with a large number of zeros at each point in time. In the Cambridge study, in any given year between 89% and 99% of males have zero convictions. Table 1a displays the 2 year intervals used for the analyses in this article. Even after the biannual grouping there are still between 83%

² ICPSR study number 8488. <http://www.icpsr.umich.edu/> includes the Cambridge data until age 24. David Farrington kindly provided us with the outcome variables until age 40 for the Cambridge sample. A summary of the non-public data is given in Table 1a.

³ Some of these boys died within the observation period. For the sake of simplicity, data for these boys are not included in our analytical illustration. The variable of interest in this context is the number of convictions per year for each of the 404 males, for whom data are available until age 31.

Table 1 Percentage (a) Conviction per age interval for the Cambridge data (b) Police contacts per age interval for the Philadelphia data (deviations from hundred are due to rounding)

(a) Convictions		10–11	12–13	14–15	16–17	18–19	20–21	22–23	24–25	26–27	28–29	30–31	32–33	34–35	36–37	38–39
Age																
0	97.1	92.7	85.6	84.2	83.2	88.8	91.4	93.3	94.6	93.3	94.1	95.0	95.0	96.5	97.8	
1	2.2	5.3	9.3	9.3	11.7	7.8	5.9	4.7	3.0	4.7	5.2	4.0	3.7	2.7	2.0	
2	0.5	0.7	3.9	2.7	2.2	1.7	1.2	1.2	2.0	1.7	0.3	0.7	0.7	0.5	0.3	
3	–	1.2	0.2	1.2	1.2	0.7	1.0	0.3	0.5	–	0.5	0.3	0.5	–	–	
4+	0.2	–	1.0	2.7	1.7	1.0	0.5	0.5	–	0.3	–	–	–	0.3	–	
	411	411	411	411	410	409	405	405	404	404	404	403	403	403	403	401
(b) Police contacts																
Age																
0	97.9	92.5	85.2	80.0	88.2	90.6	91.1	92.7								
1	1.6	5.3	8.7	10.7	7.6	6.4	5.8	5.3								
2	0.3	1.1	3.0	3.9	2.3	1.8	1.9	1.3								
3	0.1	0.6	1.3	2.2	0.9	0.7	0.6	0.5								
4+	0.1	0.5	1.7	3.2	1.0	0.5	0.6	0.3								
	13,160	13,160	13,160	13,160	13,160	13,160	13,160	13,160								

and 98% of the observations with zero convictions. The number of police contacts is equally skewed in the Philadelphia data (Table 1b). Here the percentage of zero police contacts in the bi-annual intervals ranges between 80% and 98%.

The second characteristic issue is that the aggregate age-crime curve follows the well-known pattern of increasing number of convictions throughout the subjects' teenage years and a decrease in annual conviction rates thereafter (Hirschi and Gottfredson 1983). The left panel of Fig. 1 depicts this age-crime relationship across all subjects in the Cambridge study, with the number of convictions displayed for each of the biannual observation periods used here, starting from age 10–11 up to age 38–39. However, individual trajectories do not necessarily follow this aggregate curve and, in fact, sometimes deviate from it substantially. Most males in this data set were never once convicted. But even among those with criminal careers, development of an official criminal history over time can vary considerably. The right panel of Fig. 1 shows an example of the heterogeneity in the trajectory curves. Displayed in Fig. 1 are five individual trajectories. The dot-dashed line is an example of an observation with early and high-rising conviction counts. Four convictions had been recorded in the age interval 16–17 as well as 18–19 but after age 25 there is no further conviction recorded for this case. In contrast, the dotted line shows a development with no conviction until age 31, one and respectively two counts between age 32 and 37 and zero thereafter. The dashed line is an example of a third type of trajectory with conviction counts on and off throughout the observation period.

Although the panels in Fig. 1 show the range of possible trajectories, it should be noted that the trajectories displayed in Panel 2 are somewhat “unusual” in the sense that each one appears in the data only once. The frequency of response patterns is typically not seen in articles analyzing longitudinal offense data. An example summary of pattern frequencies is given in Table 2. We will use these pattern frequencies again when we discuss model fit to data in later sections.

Table 2 summarizes patterns of conviction frequencies in biannual intervals that appear more than once in the Cambridge data. Including those males that died during the observation period, there are 108 single patterns that are omitted from this table. Most of those have a higher total number of convictions and are more likely to have convictions in two consecutive intervals.

Similar to the Cambridge data, around 60 of the subjects in the Philadelphia cohort study have no police contact throughout the entire observational period. Given the shorter time span and much larger sample there is a higher probability for multiple patterns visible

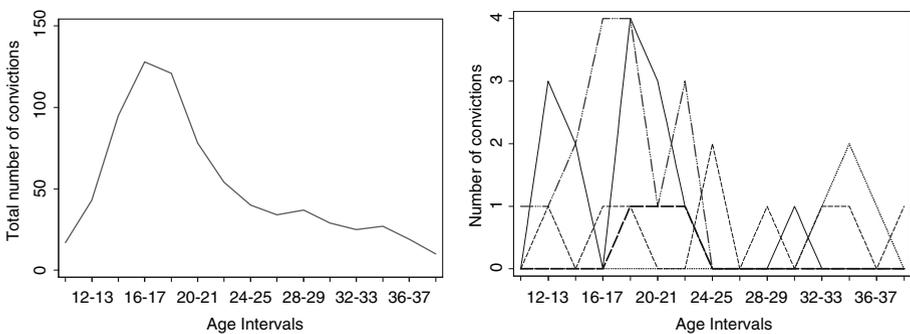


Fig. 1 Panel 1 shows the total number of convictions for all observations. Panels 2 show the developmental pathway of five men in the Cambridge data

Table 2 Patterns of convictions in biannual intervals in the Cambridge data

Pattern (Age 10–39)	<i>N</i>	%	Pattern (Age 10–39)	<i>N</i>	%
0000000000000000	245	59.61	0000001000000000	3	0.73
0000100000000000	8	1.95	0000000000001000	2	0.49
0000010000000000	6	1.46	0000000010000000	2	0.49
0010000000000000	6	1.46	0000103000000000	2	0.49
0000000000100000	5	1.22	0002000000000000	2	0.49
0001000000000000	5	1.22	0010100000000000	2	0.49
0001100000000000	5	1.22	0020000000000000	2	0.49
0000000000001000	3	0.73	0100000000000000	2	0.49
0000000100000000	3	0.73	Total displayed		73.72

Single patterns are omitted from the table

Table 3 Patterns of police contacts in biannual intervals in the Philadelphia data

Response Pattern (Age 10–25)	<i>N</i>	%	Response pattern (Age 10–25)	<i>N</i>	%
00000000	8021	60.95	00000100	201	1.53
00010000	572	4.35	01000000	181	1.38
00000010	378	2.87	00000001	141	1.07
00001000	292	2.22	00110000	117	0.89
00000010	203	1.54	00020000	107	0.81
Total displayed					77.61
Patterns with frequency $n = 1$				1,082	8.22

in Table 3. Displayed are the 10 patterns with more than 100 cases (where 100 cases are 0.76% of the data). 1,082 cases (8.22%) show a unique frequency pattern. Among those the number of police contacts throughout the entire observational period can go up to 57. However, for the vast majority of cases, 97%, the total number of police contacts between age 10 and 25 does not exceed 10, and for a cumulative total of 99% it stayed below or equal 16.

Different strategies have been used to meet these two challenges of offense trajectories. To meet the first challenge—the skewness of the count variable (number of convictions) at each point in time—the outcome variable can be analyzed in a zero-inflated Poisson (ZIP) model (Lambert 1992; Nagin and Land 1993). These models were developed for situations in which the count outcome is equal to zero more often than one would expect assuming a Poisson distribution (Hall 2000). Zero-inflated Poisson models are discussed in detail in Nagin and Land (1993) and Roeder et al. (1999), and will therefore not be explained any further. A brief description of these mixture models can be found in the Appendix. Alternatively, negative binomial models could be used to address this challenge. For simplicity we will employ ZIP to all models in this article. Our focus is on the second challenge of how to capture the heterogeneity around an overall developmental trajectory above and beyond what the zero-inflated Poisson is already capturing. More specifically, we will attempt to answer the following questions. How well can the development of a criminal career be approximated by one overall growth curve? How much individual variation is there around the overall growth curve?

How can this variation be represented? What difference does the choice of a certain modeling technique make for substantive interpretation? And, finally, how should one proceed in exploring the fit of these different models to the data?

Comparison of the Mixture Modeling Approaches

As noted above, the two dominant models used in the criminological literature to model heterogeneity of growth trajectories are growth curve models and latent class growth analysis or group-based trajectory models. We will briefly review these two models using a general latent variable framework.⁴ This framework allows for the introduction of two additional models that bridge the gap between conventional growth and group-based models, namely a parametric and non-parametric version of growth mixture models. We propose that having all four models at one's disposal will help criminological researchers understand the structure of a specific data set and will better guide modeling choices. Therefore, once the models are reviewed, we will apply each of the models in turn to the Cambridge data. A summary of the model comparisons will be provided for the Philadelphia data.

Current Modeling Strategies

Conventional Growth Modeling

In growth curve models, the joint distribution of the observed outcome variables is characterized as a function of age. Conventional growth modeling can be used to estimate the amount of variation across individuals in the growth factors (random intercepts and slopes) as well as average growth. In other words, in a conventional growth model the individual variation around the estimated average trajectory is expressed in growth factors that are allowed to vary across individuals (Raudenbush and Bryk 2002). The variation of the growth factors (random effects) is assumed to take on a normal distribution (Hedeker and Gibbons 1994). Substantively, it means that one assumes all people in the sample have the same expected criminal trajectory and the individual variation around this expected trajectory is centered on the estimated intercept and slopes for the whole sample, with symmetric deviation on both sides (e.g., some individuals start their criminal careers earlier some later, but on average they start at the estimated intercept). Conventional growth models are most often used in conjunction with covariates to “explain” the variation in the growth factors. In such cases, the people in the sample have the same expected trajectory conditioned on their covariate pattern. We will address the issue of covariates again in the discussion and focus on the simple modeling of trajectories for the present comparison.

In a simple model with no covariates other than age, the age–crime relationship is often described by a quadratic growth function (e.g., Roeder et al. 1999). In this case, the log of the Poisson rate parameter, λ , the expected value of the count part of the zero inflated Poisson model, can for each individual i and time point j be expressed as a linear

⁴ For an overview of this modeling framework, see Muthén (2002). For a step-by-step introduction to applying latent variable models to longitudinal data, see Muthén (2004). For a recent technical presentation see Muthén and Asparouhov (in press).

combination of the time-related variable a with a linear slope factor η_1 and a quadratic slope factor, η_2 ,

$$\ln(\lambda_{ij}) = \eta_{0i} + \eta_{1i}a_j + \eta_{2i}a_j^2 \quad (1)$$

Here, η_{0i} , η_{1i} and η_{2i} are random intercepts and slopes that vary across individuals. Using multilevel notation (e.g., Raudenbush and Bryk 2002) the equations estimated at Level 2 (the individual level) are:

$$\text{Level 2: } \begin{cases} \eta_{0i} = \alpha_0 + \zeta_{0i}, \\ \eta_{1i} = \alpha_1 + \zeta_{1i}, \\ \eta_{2i} = \alpha_2 + \zeta_{2i}. \end{cases} \quad (2)$$

The random effects, η_{0i} , η_{1i} and η_{2i} , can therefore also be seen as latent variables with a joint distribution that is usually assumed to be normal.

Group-based Trajectory Model/Latent Class Growth Analysis

The normality assumption for the random effects in the conventional growth model was challenged by Nagin and Land (1993). Instead Nagin and Land adopted a model by Heckman and Singer (1984) that approximates an unspecified continuous distribution of unobserved heterogeneity with a linear combination of discrete distributions (Nagin et al. 1995). That is, different groups, each with its own growth trajectory, are used to capture the overall variation.

This model can be characterized as latent class growth model. Like classic latent class analysis (see Clogg 1988), this model introduces a categorical latent variable, C ($C = 1, 2, \dots, K$), that can be viewed as a possible explanatory variable for the observed correlation between the outcome variables. Here, the parameter λ in the count part of the zero inflated Poisson is still expressed as a linear combination of the time-related variable, a , a linear slope factor, η_1 and a quadratic slope factor, η_2 ,

$$\begin{aligned} \text{Level 1: } \ln(\lambda_{ij|c_i=k}) &= \eta_{0k} + \eta_{1k}a_j + \eta_{2k}a_j^2 \\ \text{Level 2: } &\begin{cases} \eta_{0k} = \alpha_{0k}, \\ \eta_{1k} = \alpha_{1k}, \\ \eta_{2k} = \alpha_{2k}. \end{cases} \end{aligned} \quad (3)$$

However, these η s now no longer vary across individuals but across groups of individuals captured by the latent class variable. The model specifies zero within-class variance in the growth factors. The variation and covariation in the growth factors is represented exclusively by discrete mass points (corresponding to bars in a histogram). The class variable, C , replaces any distributional assumptions regarding the growth factors. The growth factors—intercept, linear and quadratic slopes of the trajectories—can vary across classes, but the model specifies zero variation within a class. This means that individuals within class k of the class variable C are treated as identical with respect to their expected developmental trajectory.

The LCGA model has been attractive for criminologists. In most of the group-based modeling applications LCGA classes have been given substantive interpretation, and LCGA is used to support distinctions among subgroups with different offending trajectories. For example, Moffitt (1993) advances a theory differentiating between a large group of adolescence-limited offenders and a small subgroup of life-course persistent offenders.

While classifying individuals in typical offender groups is appealing, there is a risk involved in the substantive interpretation of these latent class growth models. First, it is conceptually unclear whether or not the classes are used merely to model an unknown distribution of trajectories or if they instead represent substantively meaningful classes. The criminological literature is not clear on this, even across analyses of the same data set (see critique by Sampson and Laub 2005b). Second, the model assumes that the class-specific trajectory is a good representation for all members of this class. Variation around the expected trajectory within a class is assumed to be zero according to the model and therefore cannot be estimated. Third, as with all models that include latent classes, the number of classes necessary to best represent the data is often a matter of debate (e.g., D'Unger et al. 1998).

Additional Modeling Strategies

Both the conventional growth model and the latent class growth model can be seen as special cases of growth mixture models (for details see Muthén 2004 and Muthén and Asparouhov in press). The following section will briefly show the connection between these models.⁵

Growth Mixture Modeling

Similar to the conventional growth and group-based trajectory models, the quadratic growth function for a Poisson outcome in a growth mixture model can be expressed as linear combination of the time-related variable, a , a linear slope factor, η_1 and a quadratic slope factor, η_2 . As in the conventional growth model, η_{0i} , η_{1i} and η_{2i} are intercepts and slopes that may vary randomly among individuals. However, in marked contrast to conventional growth models, these random effect models can be specified for unobserved subpopulations or classes,

$$\begin{aligned} \ln(\lambda_{ij|c_i=k}) &= \eta_{0ki} + \eta_{1ki}a_j + \eta_{2ki}a_j^2 \\ \eta_{0ki} &= \alpha_{0k} + \zeta_{0ki}, \\ \eta_{1ki} &= \alpha_{1k} + \zeta_{1ki}, \\ \eta_{2ki} &= \alpha_{2k} + \zeta_{2ki}. \end{aligned} \quad (4)$$

The key differences among the classes are typically found in the fixed effects α_0 , α_1 , and α_2 , which may differ for each of the K classes of C . Intercepts and slopes may have random effects, that is, non-zero variances. If all variances in the growth factors, η_0 , η_1 , and η_2 are set to zero, a GMM model provides the same results as a latent class growth model. The LCGA can therefore be seen as a specific member of the more general class of growth

⁵ For an in depth discussion of growth mixture models see Muthén (2001a).

mixture models (Muthén 2004). Also, the conventional growth model can be seen as a growth mixture model with one class.

The growth mixture model can be estimated by maximum likelihood using an EM algorithm. For models with random effects, this involves numerical integration computation (for technical details see Muthén and Asparouhov in press). For a given solution, each individual's probability of membership in each of the classes and the individual's score on the growth factors can be estimated.

The advantage of this model is that it allows for both subpopulations, as predicted by certain criminological theories of different trajectories types, and for a variation on the theme of the trajectory type within those subpopulations. This variation can itself be predicted by covariates. However, if random effects are allowed within the classes, the GMM model relies on the normality assumption. While the assumption of normally distributed random variation among individuals might be more likely to hold within subclasses than for the whole population, it is still possible that the data-generating process has some more systematic, non-normal components.

Non-parametric Growth Mixture Modeling

As the name indicates, the non-parametric version of a growth mixture model does not rely on any distributional assumption for the random effects (Muthén and Asparouhov in press). The model is specified such that additional latent classes are estimated to capture the potentially non-normal distribution within the growth mixture classes.

Consider, for example, a growth mixture model with random intercept (no random effects for the linear and quadratic slope parameters):

$$\ln(\lambda_{ij|c_i=k}) = \eta_{0ki} + \eta_{1k}a_j + \eta_{2k}a_j^2, \text{ and } \eta_{0ki} = \alpha_{0k} + \zeta_{0ki}. \quad (5)$$

A non-parametric version of this growth mixture model would use classes to capture the ζ_{0ki} variation of the intercept within each of the K substantive classes. In this case, the distribution of the random effect for the intercept can be left unspecified and will be estimated. For estimation, the EM-algorithm can be employed. To understand the estimation procedure, it might be helpful to step back and think of numerical integration, which can be used to approximate a normal distribution. In numerical integration the integral is substituted by a finite weighted sum of mass points (nodes), similar to bars in a histogram. If one were to approximate a normal distribution with numerical integration, Gauss-Hermite quadrature can be used. In this case nodes and weights of the nodes are known and fixed. However, if an unknown distribution needs to be approximated (as it is the case here), the nodes (mass points) and weight of the nodes (masses) can be estimated. Together they provide the necessary parameters to capture the unknown distribution of the random effect.

An age–crime relationship that is described by a quadratic growth function, the random effect of the intercept within each of the K substantive classes would now no longer be captured by $\eta_{0ki} = \alpha_{0k} + \zeta_{0ki}$ but through D_k different nodes that would be called classes in the latent variable framework. The full estimation equation would be expressed with

$$\ln(\lambda_{ij|c_i=k, c_{ni}=d}) = \eta_{0kd} + \eta_{1k}a_j + \eta_{2k}a_j^2 \quad (6)$$

The subscript i on the intercept growth factor is replaced with d indicating a particular class (*node*) in the unknown distribution. The overall trajectory shape of the K substantive

classes will not change. That is, each substantive class is still defined by the same slope and quadratic term. However, in addition there will now be D_k classes within each class, k , that differ only in the estimated intercept term and that are used to capture the distribution of the intercept random effect. One can think of these non-parametric classes as shifting the growth trajectories along the y -axis. Within the non-parametric classes there are no further random effects. Hence no normality assumption is needed.

By having the slope factor means for the linear and quadratic term (η_{1k} η_{2k}) be the same across the cn -classes within each substantive class c , the assumption is made that the linear and quadratic slope growth factors are uncorrelated with the intercept growth factor. In our example this assumption is made because there was initially no random effect for the linear and quadratic slope. (Thus the only variation that the sub-classes model are for the random effect of the intercept.) This assumption is not made by the LCGA model specified in Eq. 3. In LCGA, all growth factor means (intercept, slope, and quadratic terms) can be different across classes, and all growth factor variances are set to zero. This said a latent class growth model could give a result where the estimated slope factor means (η_{1k} and η_{2k}) vary across substantive classes without being correlated with the estimated mean intercept factor.

Strategies to Decide on the Number of Classes

A common challenge for all of the latent variable models discussed here is the decision on the number of classes needed to best represent the data (see, e.g., McLachlan and Peel 2000). Objective criteria for doing so have been a matter of some controversy. The likelihood ratio test (defined as minus two times the log-likelihood of the restricted minus the log-likelihood of the unrestricted model) does not have the usual large-sample chi-square distribution due to the class probability parameter being at the border of its admissible space (Muthén 2004). In nested models (comparing a $k-1$ to a k -class model) parameter values of the k -class model are set to zero to specify the $k-1$ -class model. This model specification results in the difference of the two likelihoods not being chi-square distributed. In addition, the k -parameter space no longer has a unique maximum. Although the comparison of the log-likelihood values can indicate the appropriate number of classes, the ratio test should not be used as the sole decision criteria (Muthén 2004; Nylund et al. in press).

An alternative procedure commonly used in past criminological applications is the Bayesian Information Criteria (BIC) (see Schwarz 1978). The BIC, defined as

$$-2 \log L + p \log n, \quad (7)$$

where p is the number of parameters and n is the sample size, also makes use of the likelihood ratio and is scaled so that a small BIC value corresponds to a good model with large log-likelihood value and not too many parameters.⁶ New mixture tests were developed in the past years (for an overview see Muthén 2004), among those a bootstrap likelihood ratio-test (McLachlan and Peel 2000). The bootstrap LRT or BLRT, uses

⁶ In addition to the BIC, the Akaike Information Criteria (AIC) is sometimes used for model comparison. However, for finite mixture models, the AIC has been shown to overestimate the correct number of components (Soromenho 1994; Celeux and Soromenho 1996). The BIC on the other hand has been reported to perform well (Roeder and Wasserman 1997) and most consistently (Jedidi et al. 1997). For further details and comparisons see McLachlan and Peel (2000) as well as Nylund et al. (in press).

bootstrap samples to estimate the distribution of the log-likelihood difference test statistic. Instead of assuming the difference distribution follows a known distribution (e.g., the chi-square distribution), the BLRT empirically estimates the difference distribution. The BLRT provides a p -value that can be used to test a k - l -class model against a k -class model. In this sense the BLRT can be interpreted like the traditional likelihood ratio test, only that bootstrap sample distribution replaces the chi-square distributional assumption.

A recent simulation study by Nylund et al. (in press), in which data were generated with a known number of classes, showed that the bootstrap likelihood ratio-test (BLRT) performs better than the traditional likelihood ratio test or BIC in determining the correct number of classes. While the simulation was not carried out for GMM specifically with the Poisson or even zero-inflated Poisson outcome variables as are of interest here, the performance lags shown in the GMM setting with continuous outcomes and in the Latent Class Analysis setting were so substantial that it seems unlikely that they would surpass the performance of the bootstrap likelihood ratio-test in the present application. Nevertheless, we compare model performance using the two standard criteria (the log-likelihood test and the BIC) and supplement this information with the bootstrap likelihood test.

Fit to Data

There is not yet a formal statistical test for comparing the results of the alternative modeling approaches. The log-likelihood statistics as well as BIC used in the previous section are only one type of guide for choosing among models. Another important guide is the actual fit to the data. Model fit to data does not seem to have been commonly considered in analyses of trajectory types. We will therefore compare the different models under consideration with respect to their standardized residuals in terms of response pattern frequencies (as seen in Tables 2 and 3). We also considered the effect of single influential cases on the model fit. This is done by computing the influence statistic for each observation (Cook 1986; Liski 1991).

Application of the Four Different Modeling Strategies

In the previous section, we introduced four models for the analysis of longitudinal data. In this section, all four of these models will be applied to the analysis of the Cambridge data and the Philadelphia cohort study. All analyses are performed with maximum likelihood estimation in *Mplus* (Version 4.21, Muthén and Muthén 1998–2007). The code to reproduce our analyses is available online.⁷ For the mixture analyses of GMM, NP-GMM and LCGA, a large number of random perturbations of starting values was used to avoid local maxima.

Cambridge Data

A quadratic growth function will be specified in all models. The necessity of adding random effects on the intercept, slope and quadratic growth factors of the quadratic growth function will be explored. Further, in line with criminological theories on the existence of

⁷ <http://www.statmodel.com>

subpopulations with conceptually different developmental trajectories, the conventional growth models will be extended to growth mixture models, which allow both for separate classes and for random growth factors within those classes. Then the normality assumption within these classes will be examined and non-parametric growth mixture models will be applied. Finally, these models will be compared to latent class growth analysis in which classes represent different subpopulations with different trajectory shapes, but no variation of growth factors is allowed within these classes. For each of the models a non-technical summary will be provided.

Conventional Growth Model

This section presents the analyses of the Cambridge data using conventional growth models. In the present analysis of the Cambridge data, the outcome variable (number of convictions) is treated as zero-inflated Poisson as mentioned above.⁸ A random intercept model results in a log-likelihood value of -1481.3 with seven parameters⁹ and BIC of 3004.7 (see Table 4). According to BIC, a model that allows for not only a random-effect intercept but also for a random-effect slope fits the data even better. With these two additional parameters,¹⁰ the log-likelihood increases to -1469.6 and the BIC decreases to 2993.2. However, allowing for a random effect for the quadratic term did not lead to an improvement for BIC.

Figure 2 shows a summary of the model results. Displayed is the mean trajectory curve for the number of biannual convictions as a function of age. Not visible in the graph is the random effect variation for intercept and slope. That is, individuals are ‘allowed’ to have a different level of onset and a different slope in their trajectories. Figure 2 shows a very low average number of convictions at each time point. This is not surprising, given the large number of zeros (no convictions) for individual subjects at each point in time. The peak of the overall age–crime curve is at age 18.

Growth Mixture Model

Taking the descriptive analysis of the Cambridge data as a starting point, it seems reasonable to postulate that there exists a subgroup with no criminal career whatsoever. We therefore specified a growth mixture model (GMM) with one class being assumed to have zero values throughout the observational period. According to all model-fit indicators (log-likelihood, BIC, and BLRT) this two-class GMM model outperforms the one-class conventional growth model. The GMM model with two classes, one class with one or more convictions and one zero class (see Table 5, GMM (zip) 1 + 0), has eight

⁸ The quadratic growth function is applied to the count part as well as the zero-inflation part of the growth model. We estimated all models with and without growth structure on the zero-inflation part. The results for the models with unstructured zero-inflation part do not differ from those presented here and are therefore not listed in addition. A cubic growth parameter was specified for the conventional growth model with random intercept, but had not significant contribution. The cubic function was then not pursued any further.

⁹ Among the seven estimated parameters, three are for the means of the Poisson growth factors, one for the variance for the intercept of the Poisson growth model and three parameters for the quadratic growth model for the inflation part.

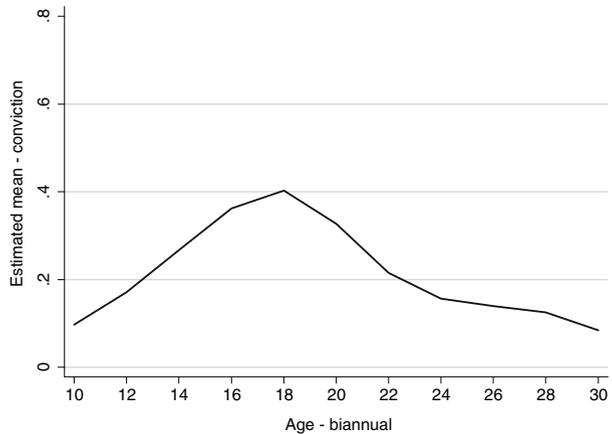
¹⁰ The two additional parameters are the variance of the slope and the covariance between slope and intercept.

Table 4 Model comparison for the Cambridge data: conventional growth model

Model	Random effect	Log-likelihood	Number of parameters	BIC
Growth (zip)	I	-1481.3	7	3004.7
Growth (zip)	I S	-1469.6	9	2993.2
Growth (zip)	I S Q	-1465.7	12	3003.5

All models are specified with a quadratic growth function. The conventional growth model differ in whether or not they allow for a random effect on the I = intercept; S = slope; or Q = quadratic growth factor of those growth function(s). Zip refers to zero-inflated Poisson

Fig. 2 Estimated mean trajectory curve of a conventional growth model with random intercept and slope for the Cambridge data



parameters. Compared to the regular growth model with a random intercept, there is now one additional parameter estimated, which is the probability for the class membership. For the GMM with one zero class the BLRT (not shown in the Table) indicates that this two-class model significantly improves the fit compared to the one class model (BLRT p -value < 0.01). The comparison of BIC (3004.7 for the one class GMM and 2994.5 for the two class GMM) and the log-likelihood values (-1481.3 for the one class GMM and -1473.3 for the two class GMM) support this decision.

As was done in the case of the general growth model above, we tested the possibility of improving model fit by allowing both random intercept and random slope parameters. This new model, with a slope variance as well as a covariance for the intercept and slope, fits the data better than the GMM with one zero class and a random intercept only. With only two more parameters, the likelihood value increases from -1473.3 to -1461.8 and BIC drops from 2994.5 to 2983.7. The presence of significant variance in the growth factors is not surprising from a substantive point of view. It seems reasonable to assume that criminal careers of a certain trajectory type would show substantial variation across individuals within this type.

Given the contribution of a random effect for the intercept and slope in the growth mixture model with one zero and one non-zero class, one might ask how this variation could best be described. Are the data best described with a model that has one non-zero class and random effects for both intercept and slope, or is it more reasonable to assume

Table 5 Model comparison for the Cambridge data: growth mixture model

Model	Classes	Random effect	Log-likelihood	Number of parameters ^f	BIC
GMM (zip)	1 + 0	I	−1473.3	8	2994.5
GMM (zip)	1 + 0	I S	−1461.8	10	2983.7
GMM (zip)	2 + 0	I	−1454.7	12	2981.5
GMM(zip)	3 + 0	I	−1450.7	16	2997.3

All models are specified with a quadratic growth function. The growth mixture model differs in whether or not they allow for an additional random effect for the I = intercept; S = slope; or Q = quadratic growth factor of those growth function(s). Zip refers to zero-inflated Poisson

variation with more than one distinct pattern for different trajectory types, as implied by the different developmental pathways described by Moffitt (1993)?

To answer these questions, we examined the improvement in model fit between the GMM with only one non-zero class and a GMM with additional non-zero classes. The GMM with one zero and two non-zero classes (in Table 5 labeled as GMM (zip) 2 + 0) has 12 parameters: Two parameters for class membership, three growth factor means for both count trajectories, one parameter for the intercept variance, and three parameters for the quadratic inflation growth part of the model.¹¹ The log-likelihood value increases to −1454.7 (from −1473.3) and BIC decreases to 2981.5 (from 2994.5) for the model with one non-zero class and random intercept (see Table 5). Adding another substantive class led to an increase in BIC and a BLRT p -value of $p = 0.1$ did not indicate a significant improvement in model fit.

The growth trajectories for the growth mixture model with the best performance are displayed in Fig. 3. For the growth mixture model one can see the estimated growth trajectories for criminal convictions for two non-zero classes, and one zero class. The dotted line at the bottom depicts the latter. The two substantive classes differ in the age at which the developmental trajectory peaks (age 16 versus 18) and they differ in the rate of decline thereafter.

The two non-zero growth factor mean curves that are displayed in Fig. 3 represent two substantial classes in terms of size. According to the model estimates, roughly 18% are in the early-peaking class and 33% are in the late-peaking class. For both classes, the model allowed the intercepts to be a random effect. The random variation of the intercept within each of the two classes is assumed to be normally distributed, an assumption that can be questioned.

One possibility for examining the appropriateness of the normality assumption is a graph of the distribution of the estimated individual values from the random intercepts. These individual values of the random effect (here the intercept factor) are often called factor scores (Lawley and Maxwell 1971). If the normality assumption holds the distribution of the individual factor scores should resemble a normal distribution. Figure 4 contains such graphs for both the early-peaking and the late-peaking classes. Plotted in Fig. 4 are histograms of the individual factor scores together with a more general kernel density estimator to smooth the function.¹²

What is noticeable in Fig. 4 is the strong skewness of the distribution in both classes. The early-peaking class also displays a slight bi-modality. Neither of these two graphs

¹¹ Note that the three inflation parameters are held equal across classes.

¹² We used an Epanechnikov kernel for the nonparametric smoothing function (Silverman, 1986).

Fig. 3 Estimated mean trajectory curves from a growth mixture model with random intercept, one zero class and two non-zero classes for the Cambridge data

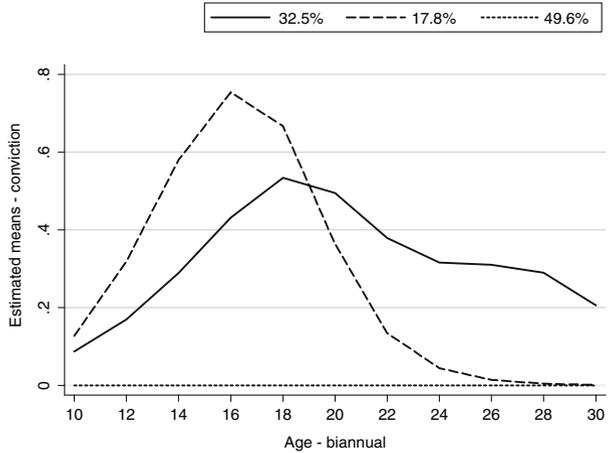
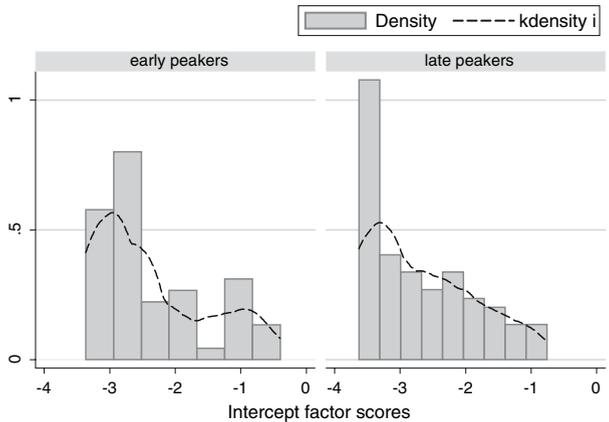


Fig. 4 Histogram and kernel density distribution of the intercept factor scores for the early- and late-peaking classes in the three-class growth mixture model for the Cambridge data



would support the assumption of symmetric normally distributed variation around the estimated mean intercept for each of these classes.

To summarize, it seems to be the case that a model that distinguishes criminal career trajectory types is preferred over the conventional growth model in which one mean trajectory is estimated for the entire sample. However, looking at the distribution of the estimated individual intercept values (i.e., factor scores) the assumption of continuous-normal varied growth within those two non-zero classes can be challenged. A non-parametric version of growth mixture models was therefore applied.

Non-parametric Growth Mixture Model

In the next step of the analysis, the assumption of normally distributed variation around the estimated intercept for both non-zero groups was relaxed. A non-parametric version of the GMM was used instead. Here, the intercept variation is represented non-parametrically through estimated support points. These can be seen as sub-classes of the two larger, substantive classes. The sub-classes differ in terms of their intercepts, but have equal slope

Table 6 Model comparison for the Cambridge data: non-parametric representation of growth mixture model

Model	Classes	Log-likelihood	Number of parameters	BIC
GMM np zip	2 (3 + 3) + 0	−1444.5	16	2985.0
GMM np zip	2 (2 + 3) + 0	−1444.4	15	2978.8
GMM np zip	2 (2 + 2) + 0	−1457.7	13	2993.0

All models are specified with a quadratic growth function. Zip refers to zero-inflated Poisson

parameters and no random effects. If the normality assumption were reasonable, then the support points (sub-classes) that are found should be symmetric and approximately equidistant from the initially estimated intercept mean.

Table 6 lists the results for three different non-parametric GMM models. All three models have two substantive classes and one zero class, just like the best fitting growth mixture model. However, now each of the substantive classes can be thought of as having two or more sub-classes (the number of sub-classes within each substantive class is indicated inside the brackets in the labels of Table 6). Each of these sub-classes will have the same overall trajectory (slope) as the larger substantive class to which it belongs. The only difference between the sub-classes is in the intercept.

In the first model reported in Table 6, three support points were used to represent the distribution of the intercepts in each of the substantive classes. However, the results of this model showed an empty class for one of the three sub-classes of the early peaking class. This indicates that only two support points are needed to represent the variation in the early-peaking class (which we found in the GMM model). This model—with two support points for intercept variation in one of the two non-zero class and three in the other—had a log-likelihood value of −1444.4 with 15 parameters and the lowest BIC value with 2978.8. Reducing the number of support points for the late-peaking class to two as was done for the early-peaking class worsens the model fit, decreasing the log-likelihood by 13 points by adding two additional parameters. Likewise, the BIC increases from 2978.8 to 2,993 (Table 6).

As mentioned earlier in the model description, numerical integration is used in the EM-algorithm. Here the integral is replaced by nodes (mass points/support points) and node weights (masses). Both node locations and weights can be estimated. Figure 5 shows the location of the support points for the distribution of the intercept factor scores for the early peakers and the late peakers. Two support points are needed for the early peakers. The height of the bars in Fig. 5 is proportionate to the class size for each of the sub-classes. Looking at both Figs. 5 and 6, one can see two early-peaking classes, those indicated by the dashed lines in Fig. 6.

About four-fifth (12.7% out of 15.5%) of the observations in the early-peaking class are estimated to have a low intercept; the average number of convictions of class members at age 16 (the peak of this trajectory) is less than 0.4. The remaining third is in the tail of the intercept distribution; the number of convictions at the peak of their trajectory is estimated to be around three. However, since the additional classes in the NP-GMM model are used to capture variation only in the intercept and not in the other parameters, both early-peaking classes are constrained to have the same slope and quadratic growth parameters. This constraint is visible in Fig. 6 in terms of the steep decline in convictions through the late teenage years and the return to zero biannual convictions by their

mid-twenties. Three support points for the intercept distribution of the late-peaking class are visible in Fig. 5. All three are displayed with the solid lines in Fig. 6.

The NP-GMM preserved the basic structure of the substantive classes found in the growth mixture model. The trajectory for the late-peaking sub-class with the lowest intercept value captures about 60% (20.2 out of 33.5) of the entire late-peaking class. The highest of the three late-peaking trajectories is the estimated mean curve for about 12% of the “late-peakers”. While the different curves imply different levels of convictions, all three trajectories show again a pattern often called “long term persistence” in the criminological literature (e.g., Roeder et al. 1999) in that criminal activity persists throughout the subject’s twenties.

Reducing the number of support points for the late-peaking class to two as was done for the early-peaking class worsens the model fit, decreasing the log-likelihood by 13 points by adding two additional parameters. Likewise, the BIC increases from 2978.8 to 2,993 (Table 6).

In summary, according to this model the data support two substantive conviction trajectories and one zero-class. There is variation within the substantive trajectories, where the level of conviction varies but not the pattern over time. Rather than capturing the variation assuming a normal distributed random effect, two or three “support points” are needed for the substantive classes.

Latent Class Growth Model

Table 7 shows a summary of the results of a Latent Class Growth Analysis of the Cambridge Study data. As stated above, a LCGA is characterized by zero variances and covariances for the growth factors. The model assumes that individuals within a class are homogeneous with respect to their development. However, unlike the non-parametric version of the growth mixture model, each class is now allowed to differ not only in the intercept factor means but also in the slope factor means. None of the classes are allowed to have random effects for any of the growth factors. The LCGA model was first estimated again having one class as a zero class.¹³ According to BIC the model with two substantive classes and one zero class clearly outperformed the model with only one substantive class (BIC 3005.4 vs. 3184.8). The model with three substantive classes and one zero class had a similar BIC value (BIC 3006.1). The 2 + 0 model (see Table 7) shows a high peaking and a low peaking class and an estimated 60% in the zero-class. In the 3 + 0 model the low peaking class “splits” into two shapes that follow the pattern we already saw in the GMM model.

Even without explicit specification of a “zero”-class a nearly zero-class is part of the resulting solution in the LCGA models for the Cambridge data. The likelihood is increased for those models. Among the LCGA models without explicit zero-class, two perform best. If model fit decision is solely based on BIC, the 4-class model performs best. The BLRT *p*-value (again not shown in the table) being <0.01 for the 5-class model indicated further significant model fit improvement, whereas adding a 6th class had no further significant improvement (BLRT *p*-value >0.05).

¹³ The inflation part was modeled in the same way as it was for the GMM. That is, three parameters are estimated and set equal across the classes.

Table 7 Model comparison for the Cambridge data: latent class growth analysis

Model	Classes	Log-Likelihood	Number of parameters	BIC
LCGA (zip)	1 + 0	−1571.4	7	3184.8
LCGA (zip)	2 + 0	−1469.7	11	3005.4
LCGA (zip)	3 + 0	−1458.0	15	3006.1
LCGA (zip)	3	−1463.7	14	3011.6
LCGA (zip)	4	−1450.0	18	3008.0
LCGA (zip)	5	−1441.0	22	3014.0
LCGA (zip)	6	−1435.2	26	3026.4

All models are specified with a quadratic growth function. Zip refers to zero-inflated Poisson

We mentioned earlier that the NP-GMM can be thought of as an LCGA model with restrictions imposed on the parameters. What is interesting to note here is that the resulting 5-class LCGA solution is very similar to the NP-GMM model results. The 5-class LCGA is characterized by one class that accounts for 68% of the sample, which has a very low intercept value and non-significant slope factors compared to all the other classes. The trajectory of this class (indicated by the dotted line in Fig. 7) is similar to the zero class in the GMM models.

The other four classes also show a familiar pattern. Two classes can be characterized by an early peak at age 16 and a sharp decline in convictions thereafter (2.7% and 11.9%) and the remaining two classes by a later peak around age 18 with some persistence through the late twenties (3.9% and 13.3%). Both of the early- and late-peaking patterns have versions with a high peak (reflecting a high number of convictions) and low peak (reflecting a low number of convictions), with a much higher proportion in the low-peak version in each case. The results of the 4-class solution look very similar. What is missing compared to the 5-class solution is the high-peaking class around age 16.

Comparing Fig. 6 and 7, the similarity between the results of the non-parametric version of GMM and the LCGA is striking. The patterns are nearly identical, which means that even if the slope and quadratic growth factors are allowed to vary among the classes (as they are in LCGA), the results are close to the result of what was called a non-parametric version of the growth mixture model.

Fit to Data

The similarity in BIC and estimated mean trajectories across the different models is underscored by looking at the standardized residuals for the Cambridge data. Similar to the model fit statistics, the standardized residuals do not point to one model with particular worse model fit.

Table 8 displays observed and estimated response patterns for the Cambridge data as well as the standardized residuals. Estimated response patterns and standardized residuals are provided for the conventional growth model, the 3-class GMM and its non-parametric version as well as two zero-class LCGA models and the 4-class and 5-class LCGA models with all trajectories estimated. In all models does the response pattern “00011000000” with one conviction at the ages 16/17 and 18/19 has a significant misfit. The pattern with

Fig. 5 Histogram intercept support points for the early and late-peaking classes in the three-class non-parametric growth mixture model for the Cambridge data

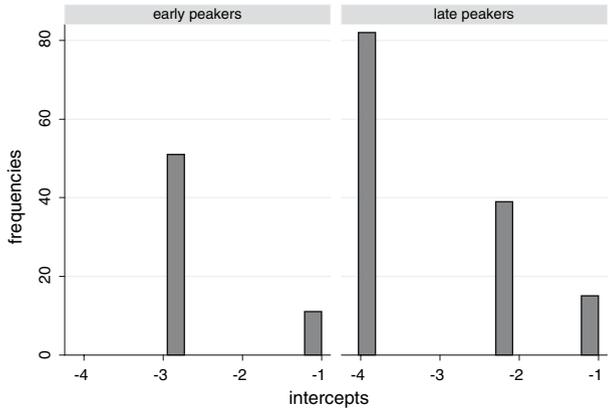


Fig. 6 Estimated mean trajectory curves in a non-parametric growth mixture model for the Cambridge data. The intercept variation is captured by separate classes with the same slope and quadratic growth parameter. Line pattern is used as a sub-class indicator. Estimated class percentages are smaller for the higher peaking sub-classes

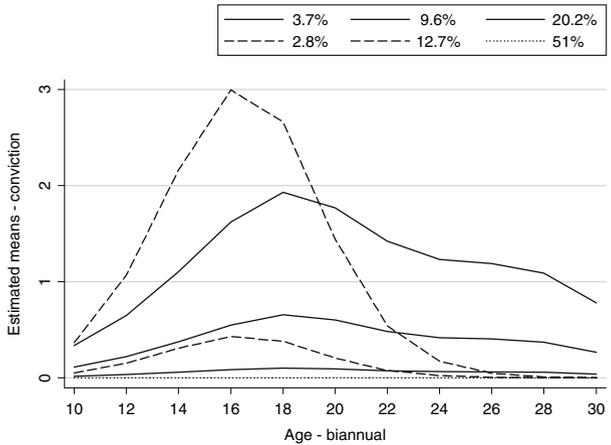
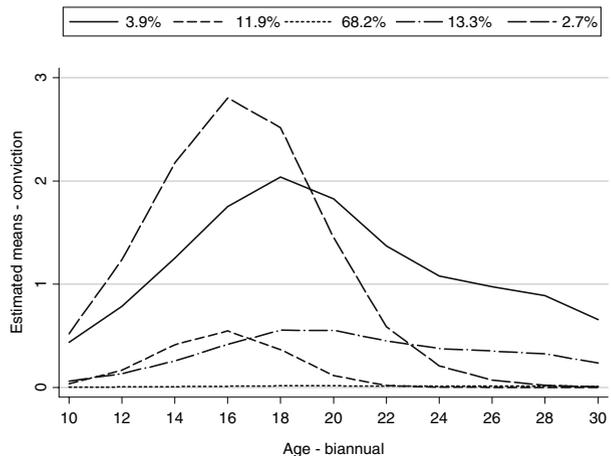


Fig. 7 Estimated mean trajectory curves in a latent class growth analysis for the Cambridge data. Line pattern is used as a class indicator



one conviction at the end of the observational period was underestimated in all five models, and significantly in GMM, NP-GMM, the 2 + 0 and 3 + 0 LCGA models.¹⁴

Summary

In sum, there are several points to note. First, all the mixture modeling alternatives indicate that at least two classes are needed in which one of the classes captures a separate developmental trajectory for young males with a very low propensity to be convicted. Second, all the mixture modeling alternatives show two trajectory ‘themes’ for the non-zero classes (see D’Unger et al. 1998; Roeder et al. 1999 for similar findings). One has an early peak with relatively quick desistance and one has a later peak with some continuation in criminal convictions throughout its members’ late thirties. Third, neither the shape of the curves nor the proportions of the classes vary substantially across the mixture modeling alternatives. The last point is especially reassuring for substantive researchers working with the Cambridge data. For one, it indicates that for the Cambridge data, the substantive conclusions on the nature of the developmental trajectories would be very similar across the different mixture modeling alternatives. Interestingly enough, this also means that although the normality assumption in GMM was violated, this violation did not influence the substantive results. One would reach essentially the same conclusions with the NP-GMM as with the regular GMM. Also, the NP-GMM yields the same conclusion as the LCGA.

We will now turn to a second model comparison using the Philadelphia cohort study. Here the model comparison leads to a different conclusion of the relative model fit. That leads us into a discussion of likely reasons for the differences with the Cambridge data application.

Modeling Results for the Philadelphia Data

From a statistical point of view, the most noticeable difference between the Cambridge data and the Philadelphia cohort study is the number of observations available for the analysis. One cannot hide the fact that the mixture models and, foremost, the random effects models are computationally demanding. It is therefore not surprising that in the past, subsets of these data had been analyzed (see D’Unger et al. 1998, 2002). Recent software and hardware improvements allow more easily for an analysis of the full data set. In the following sections we will report on the model comparison for the Philadelphia data, examine the standardized residuals and draw a comparison with the Cambridge data results.

¹⁴ What again becomes obvious here, is the small number of patterns represented by more than one person in the Cambridge data. It could be that in the case of outliers, the GMM model might have an advantage in as much as allowing for random effects can lower the effects of single influential cases. In a model that allows for variance around the growth factors, a few outliers will increase variance substantially. If the growth factors are not allowed to vary, those cases would be more likely to form a new class. Thus, we were concerned about the effect of single influential cases to the model comparison performed here. We computed the influence statistic for each observation (Cook 1986; Liski 1991). The results of the model comparison did not change after excluding influential cases (patterns).

Table 8 Observed and estimated response pattern for five best performing models for the Cambridge data

Pattern	Obs. freq.	Growth		GMM-2 + 0		NP-GMM 3(2)		LCGA-4		LCGA-5		LCGA-2 + 0		LCGA-3 + 0	
		Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.
00000000000	258	250.4	0.78	258.0	0.00	258	0.00	257.4	0.06	257.6	0.04	258	0.00	258	0.00
00001000000	9	11.1	0.63	7.7	0.48	8.2	0.30	7.7	0.47	7.8	0.44	6.51	0.99	8.70	0.10
00000100000	6	8.9	0.98	5.6	0.19	5.9	0.06	6.0	0.02	5.4	0.26	4.72	0.59	6.07	0.03
00011000000	6	1.5	3.70	1.8	3.12	2.1	2.69	1.9	3.03	2.2	2.58	1.94	2.93	2.07	2.74
00100000000	6	6.9	0.34	5.4	0.27	5.8	0.09	5.8	0.10	6.0	0.01	3.81	1.13	5.70	0.13
00010000000	5	9.7	1.53	3.1	0.90	8.2	1.12	8.0	1.05	8.5	1.22	5.76	0.32	8.17	1.12
00000000001	5	2.7	1.43	1.8	2.50	1.9	2.29	3.0	1.14	2.9	1.24	1.89	2.26	1.00	3.99
00000001000	5	4.3	0.33	2.5	1.55	2.9	1.21	3.2	1.04	3.2	1.02	2.13	1.97	2.30	1.78

Model Comparison

For the modeling comparison with the Philadelphia cohort data we use the full set of 13,160 observations (excluding 34 individuals with more than 10 criminal offenses in any given year, the analysis are based on 13,126).¹⁵ Like in the Cambridge data analyses a quadratic model was applied to the data. Table 9 shows the modeling results for the conventional growth model, the growth mixture model with random effects and in its nonparametric version as well as results from the latent class growth analysis.

Unlike for the Cambridge data, here the growth mixture models included random effects for all three terms; the intercept, linear slope, and quadratic slope. Allowing the effects to be random for the linear and the quadratic slope improved the model fit. In all of the models displayed here, no structure was applied to the inflation part of the growth model.¹⁶ Compared to the conventional growth model, BIC improved by adding classes to the random effect models. A 3-class mixture GMM model showed clear improvement in BIC compared to the conventional growth model. The 4th class added little improvement compared to the 3-class GMM. Figure 8 shows the results for the 3-class GMM model. The largest class (64.6%) has an estimated mean trajectory almost flat at zero. The second estimated mean trajectory shows a peak in teenage years with a clear decline after age 17, flattening out around zero, similar to the largest class at age 20. The third trajectory type found for the Philadelphia data peaks at age 16/17 and shows continuing police contacts until the end of the observational period at age 25.

For the LCGA models, the BIC kept improving up to as many as eight classes. However, with three classes, the GMM gives a better BIC value than the 8-class LCGA model. Finding a dip in BIC values can be difficult for data with large number of observations. Substantive considerations were taken into account by D'Unger et al. (1998, 2002) as well as Loughran and Nagin (2006) to decide on the number of classes. Loughran and Nagin (2006, pp. 255) give good substantive reasons for their choice of a 4-class model. The resulting trajectories are comparable to our LCGA modeling results. With the same number of parameters the non-parametric GMM has a better BIC value than the 4-class LCGA. Looking at the likelihood values, the 3-class GMM has a better log-likelihood by 200 points with two parameters less than the 5-class LCGA.

Fit to Data

Tables 10 and 11 show the model fit to the 10 most frequent patterns, comprising 78% of the Philadelphia data. Both tables show the estimated frequencies for each of the models. Table 10 for the conventional growth model, the three GMM models and the non-parametric GMM model. Table 11 shows the estimated frequencies for five different LCGA models. The column next to each estimated frequency has the standardized residuals for this particular pattern.¹⁷ Residuals significant at the 5% level are highlighted. To give an example, in the first column of Table 10 the response pattern “00001000”, with one police contact between age 18 and 19 and no police contact before or thereafter, appears 292

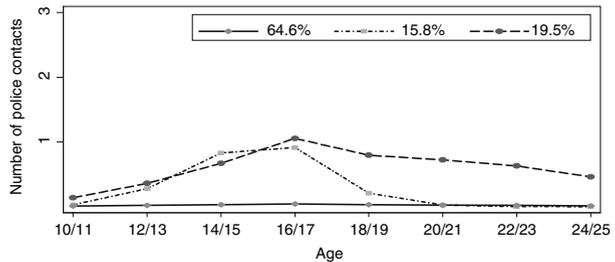
¹⁵ A similar strategy was employed by Loughran and Nagin (2006) in their analysis of the full data set.

¹⁶ Unlike in the Cambridge analysis, the model here has one inflation parameter per time point, held equal across classes.

¹⁷ For LCGA and the non-parametric GMM those can be computed by hand. For GMM with random effects, numerical integration was used.

Table 9 Summary of the modeling results for the Philadelphia cohort study

Model	Classes	Random effect	Log-likelihood	Number of parameters	BIC
Conventional Growth (zip)	1	I S Q	-40,606	17	81,373
GMM (zip)	3	I S Q	-40,283	25	80,803
GMM (zip)	4	I S Q	-40,237	29	80,748
GMM np zip	3(2)		-40,458	23	81,133
LCGA (zip)	4		-40,643	23	81,503
LCGA (zip)	5		-40,483	27	81,222

Fig. 8 Estimated mean trajectories for the Philadelphia cohort study. Results from the 3-class GMM

times in the data. The conventional growth model would overestimate its appearance by 60 respondents. The conventional growth model has in total five significant residuals. With two additional classes and random effects for all growth factors the 3-class GMM has only one significant residual and the 4-class GMM has none.

In contrast, the 4-class LCGA model has four significant residuals and the 5-class model has three. An 8-class LCGA is needed to reduce this to one significant residual. Not surprisingly did we see a better BIC value for the 3- and 4-class GMM compared to the LCGA models with 8 or fewer classes. The residual analysis supports the earlier notion of a better model fit to the data for the GMM compared to NPGMM and LCGA.

Comparison to the Cambridge Data

Both the comparison of BIC across models as well as the standardized residuals supported the GMM choice for the Philadelphia data. The 3-class and 4-class GMM were estimated with random effects for the intercept, the linear slope and the quadratic slope parameters. The large number of observations in the Philadelphia data compared to the Cambridge data increased the power to detect differences in model fit between the different models presented here. For the Cambridge data, the lack of power in detecting significant differences in model fit turned out to not be problematic. The resulting trajectory curves showed comparable location, form, and size across the different models.¹⁸ The question is would the same be true for the Philadelphia data? Would the model comparison have led to a

¹⁸ Going through the different modeling steps, the non-parametric GMM nicely bridged the results from LCGA and GMM for the Cambridge data. Without having yet looked at covariates or predictive power, the case can be made that there are only three overall patterns in the Cambridge data with high and low variations on the themes.

Table 10 Observed and estimated response pattern for the conventional growth model, 2-, 3-, 4-class GMM as well as the non-parametric GMM with three substantive classes and two non-parametric classes

Response pattern	Obs. freq.	Growth		GMM 2		GMM 3		GMM 4		NPGMM	
		Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.	Est. freq.	Std. res.
00000000	8,021	7,851	3.03	7,871	2.68	8,015	0.11	8,013	0.15	7978.4	0.76
00010000	572	670	3.88	714	5.49	577	0.20	582	0.44	559.3	0.55
00100000	378	431	2.61	472	4.41	371	0.37	380	0.13	377.8	0.01
00001000	292	352	3.25	312	1.13	300	0.44	294	0.11	303.7	0.68
00000010	203	232	1.92	192	0.79	197	0.42	198	0.33	189.9	0.95
000000100	201	265	3.98	202	0.07	226	1.68	220	1.31	231.4	2.02
01000000	181	172	0.71	160	1.67	156	2.01	159	1.72	179.0	0.15
000000001	141	157	1.29	168	2.09	147	0.53	154	1.06	129.2	1.04
00110000	117	110	0.69	126	0.82	116	0.09	118	0.06	88.6	3.03
00020000	107	94	1.37	105	0.16	105	0.22	112	0.48	89.1	1.90

Full data for the Philadelphia cohort study

Table 11 Observed and estimated response pattern for 4-, 5-, 6-, 7-, and 8-class LCGA

Response pattern	Obs. Freq.	LCGA-4		LCGA-5		LCGA-6		LCGA-7		LCGA-8	
		Est. freq.	Std. res.								
00000000	8,021	7,965	1.00	8,003	0.32	8,009	0.21	8,020	0.03	8,021	0.00
00010000	572	581	0.37	565	0.31	561	0.48	560	0.50	547	1.10
00100000	378	382	0.22	377	0.05	374	0.22	376	0.11	384	0.32
00001000	292	329	2.07	313	1.17	307	0.85	300	0.46	304	0.71
00000010	203	212	0.62	198	0.38	198	0.32	190	0.92	190	0.97
00000100	201	257	3.54	238	2.45	239	2.48	228	1.78	228	1.81
01000000	181	177	0.30	176	0.39	174	0.56	173	0.62	173	0.62
00000001	141	145	0.36	132	0.78	132	0.76	128	1.12	128	1.11
00110000	117	81	4.08	82	3.88	88	3.15	92	2.61	92	2.60
00020000	107	85	2.40	87	2.14	87	2.15	87	2.14	93	1.49

Full data for the Philadelphia cohort study

different conclusion with a smaller sample size? And would the inference drawn from the resulting model estimates have been comparable? Or would the significant random effects for the linear and especially the quadratic slope have led to different conclusions?

To answer these questions, we selected a random sample of $n = 500$ of the Philadelphia cohort study. This sample size is comparable to that of the Cambridge data. After reducing the sample size, the random effect for the quadratic growth factor is no longer significant and does not provide any improvement in model fit.¹⁹ This is true for the conventional growth model as well as for growth mixture models with random effects. Second, unlike in the full data, the growth mixture model with three classes has a better BIC value (BIC = 2,945) than the 4-class model (BIC = 2,966). Interestingly, the conventional growth model with random effects for the intercept and slope has a better BIC (=2,937) than the growth mixture models for this restricted data set. The 4-class LCGA has a better BIC than both the 3- and the 5-class LCGA models (BIC = 2,978 compared to BIC = 3,000 and BIC = 2,982 respectively), but not as good as any of the growth models with random effects. Third, the log-likelihood values of all these models are very close and differences are comparable in size to what we have seen in the modeling results for the Cambridge data. Based solely on the subsample of $n = 500$, it would be difficult to make a definite judgment on the superiority of one model over the other based on the statistical measures available. Had one only used LCGA, the decision would have been to settle on a 4-class model. Looking at Fig. 9 however, one can see that the resulting mean trajectories differ in shape and size for two of the LCGA classes compared to the 3-class GMM model (which was found to fit best for the full data set). The two estimated trajectories for the largest class and the low peaking class (solid line in both graphs) are very similar in both models. The higher peaking GMM trajectory type seems to subsume the variation of the additional two estimated trajectory types from the 4-class LCGA.

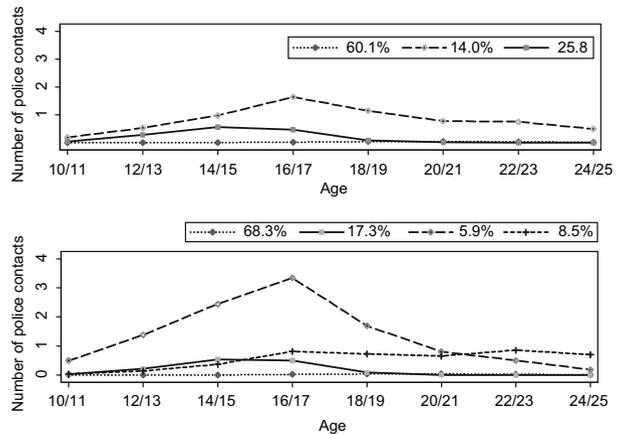
It appears that the combination of random intercept and random slope for the GMM models makes the results across the modeling types less similar. A small sample size will also make it difficult to evaluate significant differences in model fit. For the $n = 500$ Philadelphia data—just looking at BIC for model fit—a conventional growth model with random intercept and slope would have been the choice. The analysis for the full Philadelphia cohort study showed however, that the growth mixture model had an improved model fit. In situations with a sample of this size, one can only hope (or plan before the data collection) that covariates are available that allow for an examination of predictive validity and correlations to antecedents.

Summary and Discussion

This article illustrates how four different modeling approaches can be used to analyze growth trajectories. The approaches considered included both the two dominant models used in the criminological literature—growth curve models and latent class growth analysis—and parametric and non-parametric versions of growth mixture models. These models were applied to two data sets. The application of all of these models within a general latent variable framework allowed for a straightforward comparison of the models, while the use of the same data permits direct comparison of the modeling results in terms of the models' ability to capture the heterogeneity in trajectories. We demonstrated that

¹⁹ BIC with random effect for the quadratic slope factor was 2951.3 compared to 2936.5 with just random effects on the intercept and linear slope.

Fig. 9 Estimated mean trajectories for the $n = 500$ subsample of the Philadelphia cohort study. Upper panel 3-class GMM, lower panel 4-class LCGA model



researchers do not have to make a priori decisions on whether to assume continuously varied growth or to rely entirely on substantive classes to capture the variation in growth. A growth mixture model where random effects are allowed within classes can be an alternative. Or, if the normality assumption is questioned, a non-parametric growth mixture model version can be considered before the variation is modeled entirely group-based.

For the Cambridge data, substantive researchers can take heart in the essentially equivalent results of the models compared here. For the Philadelphia cohort study, the results differ across the models and substantively different conclusions are likely to be drawn if only one of the different modeling approaches would be used (especially with a reduced sample size). The model comparison in this article was largely based on BIC as well as an examination of the standardized residuals to address model fit to data. For substantive researchers this model comparison should only be the beginning. Ultimately, model choices should be supported if not guided by arguments related to substantive theory, auxiliary information, predictive validity, and practical usefulness. Especially in cases where power is limited due to sample size. To this point we want to emphasize three issues to be considered in applying any of the models discussed in this article.

First, an important part of mixture models is the prediction of class membership probabilities from covariates. This gives the profiles of the individuals in the classes. If theories differentially relate auxiliary information in the form of covariates of class membership and growth factors, those should be included in the set of covariates to correctly specify the model, find the proper number of classes, and correctly estimate class proportions and class membership (Muthén 2002). The fact that the “unconditional model” without covariates is not always suitable for finding the number of classes has not been fully appreciated (Muthén 2004).

Second, similar to the examination of covariates, the predictive power of different trajectory types for later outcomes should be considered in the modeling. For example, if the fit criteria used here point to a mixture model, but all classes have the same predictive power for later outcomes and are predicted by the same covariates in the same way, there is strong support for interpreting the mixture components as non-parametric versions of a general growth model. For the Cambridge data, the question is: Do the sub-classes that are used for the non-parametric representation of the variation in the growth factor intercept show different effects on the number of convictions in later time points? In the Cambridge example this question is especially interesting for the NP-GMM and the LCGA solution. It

would be harder to argue that these sub-classes are solely used to represent a non-normal distribution of the growth factor(s) in the GMM model if they show different predictive power for convictions at for example age 38–40. A model including the distal outcome (convictions at age 38–40) showed indications of such a situation. In the non-parametric model, one of the three sub-classes (represented by the solid line with 9.6% class proportion in Fig. 6) had a higher probability of conviction in their late 30s than those in the other two sub-classes.²⁰ In the LCGA model, only one class has a probability that is significantly different from zero for convictions in the respondents' late 30's, and that is one of the two late-peaking classes (represented by the dot-dash pattern in Fig. 7).

However, one should note that being able to predict a distal outcome from trajectory class memberships does not necessarily constitute evidence for multiple classes (Muthén 2003). For example, if data have been generated by a conventional single-class growth model where increasing growth factor intercept and slope values gives an increasing probability of the distal outcome, a mixture model (here GMM, NP-GMM or LCGA) might point to a 2-class solution with a high and a low class where the high class has a higher distal outcome probability. The same can be said for the high and low classes that split the classes found with the GMM model for the Cambridge data. When statistical evidence is lacking, here too, substantive considerations are key in the analysis.

Finally, researchers should also keep in mind that a model comparison can lead to quite different results for different data, and differently scaled outcome variables. Modeling comparisons for ordered categorical and continuous outcomes can be found in Muthén (2001a, 2001b, 2004). More research needs to be done on binary outcomes and counts. Hopefully this article gives guidance and inspiration on how such modeling comparisons can be approached.

Acknowledgments We like to acknowledge everyone who discussed this article with us during the last 5 years at various meetings and conferences. In particular we thank Tihomir Asparouhov for ongoing discussions that shaped our perspective on this article. Shawn Bushway, Booil Jo, John Laub, Katherine Masyn, Daniel Nagin, Paul Nieuwebeerta and three anonymous reviewers provided critical comments to earlier versions of this manuscript that we gladly took into account. Michael Lemay was of great help in data preparation and analysis. The work on this article was partially supported by grant K02 AA 00230 from National Institute on Alcohol Abuse and Alcoholism.

Appendix

Zero-inflated Poisson Model

Note that the ZIP model is already a special case of a finite mixture model with two classes. Treating the count outcome variable as zero-inflated at each time point means that a probability is estimated for the observation to be either in the “zero-class” or not. For the zero class a zero count occurs with probability one. For the non-zero class, the probability of a conviction is expressed with a Poisson process.

The interesting feature for the ZIP, or its expression as a two-class model, is that the probability of being in the zero class can be modeled by covariates that are different from

²⁰ The non-parametric GMM in which the three sub-classes are constrained to have equal probability for later conviction shows a slight decrease in fit. The log-likelihood value for the restricted model is -1467.8 with 19 parameters compared to -1462.0 with 21 parameters in the unrestricted model.

those that predict the counts for the Poisson class. The same is true when allowing for a zero class in the growth trajectory modeling.

More formally, for the present application this model can be represented as follows: At each individual time point a count outcome variable U_{it} (the number of conviction at each time point t for individual i) is distributed as ZIP (Roeder et al. 1999).

$$U_{it} \sim \begin{cases} 0 & \text{with probability } \rho_{ij} \\ \text{Poisson}(\lambda_{it}) & \text{with probability } 1 - \rho_{ij} \end{cases}$$

The parameters ρ_{it} and λ_{it} can be represented with $\text{logit}(\rho_{it}) = \log[\rho_{it}/1 - \rho_{it}] = \mathbf{X}_{it}\gamma_t$ and $\log(\lambda_{it}) = \mathbf{X}_{it}\beta_t$.

Notice that the mixture model within the zero-inflated Poisson is a mixture at each time point. The mixture models we discuss in the different growth models are mixtures of different growth trajectories (across all time points).

References

- Blokland A, Nagin DS, Nieuwebeerta P (2005) Life span offending trajectories of a Dutch conviction cohort. *Criminology* 43:919–953
- Celeux G, Soromenho G (1996) An entropy criterion for assessing the number of clusters in a mixture model. *J Class* 13:195–212
- Clogg CC (1988) Latent class models for measuring. In: Langeheine R, Rost J (eds) *Latent trait and latent class models*. Plenum, New York
- Cook RD (1986) Assessment of local influence (with discussion). *J.R. Stat Soc, B* 48:133–169
- D'Unger AV, Land KC, McCall PL, Nagin DS (1998) How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses. *Am J Sociol* 103:1593–1630
- D'Unger AV, Land KC, McCall PL (2002) Sex differences in age patterns of delinquent/criminal careers: Results from Poisson latent class analyses of the Philadelphia cohort study. *J Quant Criminol* 18:349–375
- Elliot D (1985) *National Youth Survey 1976–1980: Wave I-V*. Behavioral Research Institute, Inter-University Consortium for Political and Social Research, Ann Arbor, Michigan
- Farrington DP, West DJ (1990) The Cambridge study in delinquent development: A long-term follow-up of 411 London males. In: Kerner HJ, Kaiser G (eds) *Kriminalität: Persönlichkeit, Lebensgeschichte und Verhalten*. Festschrift für Hans Göppinger zum 70. Springer, Geburtstag
- Farrington DP, West DJ (1993) Criminal, penal and life histories of chronic offenders: risk and protective factors and early identification. *Crim Behav Ment Health* 3:492–523
- Hall DB (2000) Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56:1030–1039
- Heckman J, Singer B (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52:271–320
- Hedeker D, Gibbons RD (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics* 50:933–944
- Hirschi T, Gottfredson M (1983) Age and the explanation of crime. *Am J Sociol* 89:552–584
- Jedidi K, Jagpal HS, DeSarbo WS (1997) Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Mark Sci* 16:39–59
- Jones B, Nagin D, Roeder K (2001) A SAS procedure based on mixture models for estimating developmental trajectories. *Sociol Methods Res* 29:374–393
- Lambert D (1992) Zero-inflated Poisson regression with application to defects in manufacturing. *Technometrics* 34:1–14
- Laub J, Sampson RJ (2003) *Shared beginnings, divergent lives. Delinquent boys to age 70*. Harvard University Press, Cambridge
- Lawley DN, Maxwell AE (1971) *Factor analysis as a statistical method*, 2nd edn. American Elsevier, New York
- Liski EP (1991) Detecting influential measurements in a growth curve model. *Biometrics* 47:659–668
- Loughran T, Nagin DS (2006) Finite sample effects in group-based trajectory models. *Sociol Methods Res* 35:250–278

- McLachlan G, Peel D (2000) Finite mixture models. John Wiley, New York
- Moffitt T (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev* 100:674–701
- Muthén B (2001a) Latent variable mixture modeling. In: Marcoulides GA, Schumacker RE (eds) *New developments and techniques in structural equation modeling*. Lawrence Erlbaum Associates, pp 1–33
- Muthén B (2001b) Second-generation structural equation modeling with a combination of categorical and continuous latent variables: new opportunities for latent class/latent growth modeling. In: Collins LM, Sayer A (eds) *New methods for the analysis of change*. APA, Washington, D.C., pp 291–322
- Muthén B (2002) Beyond SEM: general latent variable modeling. *Behaviormetrika* 29:81–117
- Muthén B (2003) Statistical and substantive checking in growth mixture modeling: comment on Bauer and Curran. *Psychol Methods* 8:369–377
- Muthén B (2004) Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. In: Kaplan D (ed) *Handbook of quantitative methodology for the social sciences*. Sage, Newbury Park, pp 345–368
- Muthén B, Asparouhov T (in press) Growth mixture modeling: analysis with non-Gaussian random effects. Draft available at: <http://www.statmodel.com/download/ChapmanHall06V24.pdf> Forthcoming in: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) *Advances in longitudinal data analysis*. Chapman & Hall/CRC Press
- Muthén LK, Muthén BO (1998–2007). *Mplus user's guide*, 4th edn. Muthén & Muthén, Los Angeles, CA
- Nagin DS (1999) Analyzing developmental trajectories: a semi-parametric, group-based approach. *Psychol Methods* 4:139–157
- Nagin DS, Farrington DP, Moffitt TE (1995) Life-course trajectories of different types of offenders. *Criminology* 33:111–137
- Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology* 31:327–362
- Nieuwebeerta P, Blokland A (2003) Criminal careers in adult Dutch offenders (Codebook and Documentation). NSCR, Leiden
- Nylund KL, Asparouhov T, Muthén B (in press) Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Draft available at: http://www.statmodel.com/download/LCA_tech11_nylund_v83.pdf To appear in: *Structural Equation Modeling: A Multidisciplinary Journal*
- Piquero AR (2007) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman A (ed) *The long view of crime: a synthesis of longitudinal research*. Springer, New York
- Piquero AR, Buka SL (2002) Linking juvenile and adult pattern of criminal activity in the providence cohort to the National Collaborative Perinatal Project. *J Crim Justice* 30:259–272
- Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm: background and recent developments. *Crime and Justice: A Rev Res* 30:359–506
- Raudenbush SW, Bryk AS (2002) *Hierarchical linear models: applications and data analysis methods*, 2nd edn. Sage Publications, Newbury Park
- Raudenbush SW (2005) How do we study “what happens next”? *Ann Am Acad Pol Soc Sci* 602:131–144
- Roeder K, Lynch K, Nagin D (1999) Modeling uncertainty in latent class membership: a case study in criminology. *J Am Stat Assoc* 94:766–776
- Roeder K, Wasserman L (1997) Practical Bayesian density estimation using mixtures of normals. *J Am Stat Assoc* 92:894–902
- Sampson RJ, Laub JH (2005a) A life-course view of the development of crime. *Ann Am Acad Pol Soc Sci* 602:12–45
- Sampson RJ, Laub JH (2005b) Seductions of methods: rejoinder to Nagin and Tremblay’s “developmental trajectory groups: fact or fiction?”. *Criminology* 43:905–913
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Silverman BW (1986) *Density estimation for statistics and data analysis*. Chapman and Hall, London
- Soromenho G (1994) Comparing approaches for testing the number of components in a finite mixture model. *Comput Stat* 9:65–78
- Tracy P, Wolfgang ME, Figlio RM (1990). *Delinquency careers in two birth cohorts*. Plenum Press, New York