

## Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach

Daniel S. Nagin

Carnegie Mellon University

A developmental trajectory describes the course of a behavior over age or time. A group-based method for identifying distinctive groups of individual trajectories within the population and for profiling the characteristics of group members is demonstrated. Such clusters might include groups of “increasers,” “decreasers,” and “no changers.” Suitably defined probability distributions are used to handle 3 data types—count, binary, and psychometric scale data. Four capabilities are demonstrated: (a) the capability to identify rather than assume distinctive groups of trajectories, (b) the capability to estimate the proportion of the population following each such trajectory group, (c) the capability to relate group membership probability to individual characteristics and circumstances, and (d) the capability to use the group membership probabilities for various other purposes such as creating profiles of group members.

Over the past decade, major advances have been made in methodology for analyzing individual-level developmental trajectories. The two main branches of methodology are hierarchical modeling (Bryk & Raudenbush, 1987, 1992; Goldstein, 1995) and latent curve analysis (McArdle & Epstein, 1987; Meredith & Tisak, 1990; Muthen, 1989; Willett & Sayer, 1994). These advanced methods allow researchers to move beyond the use of ad hoc categorization procedures for constructing developmental trajectories.

Although these two classes of methodology differ in very important respects, they also have important commonalities (MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Muthen & Curran, 1997; Raudenbush, in press; Willett & Sayer, 1994). For the purposes of this article, one is key: Both model the

unconditional and conditional population distribution of growth curves based on continuous distribution functions. Unconditional models estimate two key features of the population distribution of growth curve parameters—their mean and covariance structure. The former defines average growth within the population, and the latter calibrates the variances of growth throughout the population. The conditional models are designed to explain this variability by relating growth parameters to one or more explanatory variables.

This article demonstrates a distinct semiparametric, group-based approach for modeling developmental trajectories. The modeling strategy is intended to provide a flexible and easily applied approach for identifying distinctive clusters of individual trajectories within the population and for profiling the characteristics of individuals within the clusters. Using mixtures of suitably defined probability distributions, the method identifies distinctive groups of developmental trajectories within the population (Land & Nagin, 1996; Nagin, Farrington, & Moffitt, 1995; Nagin & Land, 1993; Nagin & Tremblay, in press; Roeder, Lynch, & Nagin, in press). Thus, whereas the hierarchical and latent curve methodologies model population variability in growth with multivariate continuous distribution functions, the group-based approach uses a multinomial modeling strategy and is designed to

---

This work was supported by National Science Foundation Grant SBR-9511412 and by the National Consortium for Violence Research. I thank Robert Brame, Lisa Broidy, Patrick Curran, Kenneth Dodge, Laura Dugan, Anne Garvin, Don Lynam, Steve Raudenbush, Kathryn Roeder, Robert Siegler, and Larry Wasserman for helpful comments on previous versions of this article.

Correspondence concerning this article should be addressed to Daniel S. Nagin, H. John Heinz III School of Public Policy & Management, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

identify relatively homogeneous clusters of developmental trajectories.

The group-based modeling strategy demonstrated in this article is presaged in the work of Rindskopf (1990). Rindskopf's methodology is fully nonparametric. It is designed for analysis of repeated measurements of dichotomous response data. The repeated measurements may come in the form of responses to a series of test items or, alternatively, in the form of panel data. Rindskopf's method is designed to identify distinct groups of response sequences across sampled individuals. The method described here expands on Rindskopf's innovative work in several important respects—it increases the variety of response variables to which a group-based modeling strategy can be applied, it provides a basis for linking group membership probability to relevant individual-level characteristics, and it describes a formal approach for determining the optimal number of groups.

### Rationale for a Group-Based Modeling Strategy

Before discussing the technical features of the model, I briefly speak to the underlying rationale for a group-based modeling strategy. There is a long tradition in psychology of group-based theorizing about development. Examples include theories of personality development (Caspi, 1998), drug use (Kandel, 1975), learning (Holyoak & Spellman, 1993), language and conceptual development (Markman, 1989), development of prosocial behaviors such as altruism, and development of antisocial behaviors such as delinquency (Loeber, 1991; Moffitt, 1993; Patterson, DeBaryshe, & Ramsey, 1989). This group-based approach is well suited to analyzing questions about developmental trajectories that are inherently categorical—do certain types of people tend to have distinctive developmental trajectories?

For example, Moffitt (1993) and Patterson et al. (1989) have argued that early onset of delinquency is a distinctive feature of the development trajectories of chronically criminal and antisocial adults but not of individuals whose offending is limited to their adolescence. The group-based modeling strategy can test whether the developmental trajectories predicted by the theories are actually present in the population. It can also be used to test key predictions such as whether the chronic trajectory is typified by an earlier onset of delinquent behavior than the adolescent limited trajectory. In contrast, because hierarchical and latent

curve modeling assume a continuous distribution of trajectories within the population, these methods are not designed to identify distinct clusters of trajectories. Rather they are designed to describe how patterns of growth vary continuously throughout the population. As a result, it is awkward to use these methods to address research questions that contrast distinct categories of developmental trajectories.

The group-based modeling approach assumes that the population is composed of a mixture of distinct groups defined by their developmental trajectories. The assumption that the population is composed of distinct groups is not likely literally correct. Unlike biological or physical phenomena, in which populations may be composed of actually distinct groups such as different types of animal or plant species, population differences in developmental trajectories of behavior are unlikely to reflect such bright-line differences (although biology has a long tradition of debates concerning classification; see Appel, 1987).

To be sure, taxonomic theories predict different trajectories of development across subpopulations, but the purpose of such taxonomies is generally to draw attention to differences in the causes and consequences of different developmental trajectories within the population rather than to suggest that the population is composed of literally distinct groups. In this respect, developmental taxonomies are analogous to prototypal diagnostic classifications in clinical psychology. Unlike classical classifications, prototypal classifications acknowledge fuzziness in classification, employ polythetic diagnostic criteria, and recognize within-group variation in the prevalence and severity of symptoms (Cantor & Genero, 1986; Widiger & Frances, 1985).

The group-based modeling strategy is prototypal in design and provides a methodological complement to theories that predict prototypal developmental etiologies and trajectories within the population. The modeling strategy explicitly recognizes uncertainty in group membership, allows an examination of the impact of multiple factors on probability of group membership, and anoints no set of factors as necessary and sufficient in determining group membership. With this approach the basic elements of taxonomic theories can be directly tested: Are the relatively homogeneous developmental trajectories and etiologies predicted by the theory actually present in the population?

In so doing, the method provides an alternative to using assignment rules based on subjective categori-

zation criteria to construct categories of developmental trajectories. Although such assignment rules are generally reasonable, there are limitations and pitfalls attendant to their use. One is that the existence of the various developmental trajectories that underlie the taxonomic theory cannot be tested; they must be assumed a priori. A related pitfall of constructing groups with subjective classification procedures is *overfitting*—creation of trajectory groups that reflect only random variation. Second, *ex ante* specified rules provide no basis for calibrating the precision of individual classifications to the various groups that compose the taxonomy. The semiparametric, group-based method presented in this article avoids each of these limitations. It provides a formal basis for determining the number of groups that best fits the data and also provides an explicit metric, the posterior probability of group membership, for evaluating the precision of group assignments.

#### Data and Three Illustrative Examples

Throughout, key methodological points are illustrated with findings from the analysis of two well-known longitudinal studies. One is the Cambridge Study of Delinquent Development, which tracked a sample of 411 British males from a working-class area of London. Data collection began in 1961–62, when most of the boys were age 8. Criminal involvement is measured by conviction for criminal offenses and is available for all individuals in the sample through age 32, with the exception of the 8 individuals who died prior to this age. Between ages 10 and 32 a wealth of data was assembled on each individual's psychological makeup, family circumstances including parental behaviors, and performance in school and work. For a complete discussion of the data set, see Farrington and West (1990).

The other data set is from a prospective longitudinal study based in Montreal, Canada. This study has been tracking 1,037 White males of French ancestry. Participants were selected in 1984 from kindergarten classes in low-socioeconomic-status Montreal neighborhoods. Following the assessment at age 6, the boys and other informants were interviewed annually from ages 10 to 18. Like the Cambridge study, assessments were made on a wide range of factors. Among these is physical aggression, which was measured on the basis of teacher ratings using the Social Behavior Questionnaire (Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987). The boys were rated on three items: kicking, biting, and hitting other children;

fighting with other children; and bullying and intimidating other children. See Tremblay et al. for further details on this study.

Figures 1, 2, and 3 show the results of trajectory analyses based on these two data sets. The software used to estimate these models is a customized SAS procedure that was developed with the SAS product SAS/TOOLKIT. The procedure, which is currently available for the PC platform only, is a program written in the C programming language and is designed to interface with the SAS system to perform model fitting. It accommodates missing data. Thus, individuals with incomplete assessment histories do not have to be dropped from the analysis. Also, time between assessment periods does not have to be equally spaced, and assessment periods do not have to be identical across participants. The software, which is easily used and incorporated into the standard SAS package, is available on request and is described in detail in Jones, Nagin, and Roeder (1998).

Figure 1 displays the results of the analysis of the Cambridge data. The response variable in this analysis is a count, the number of convictions for each 2-year period from ages 10 to 31 years. The solid lines represent actual behavior, and the dashed lines represent predicted behavior.<sup>1</sup> Three groups were identified: A group called "the never convicted" was composed of respondents who, but for a few individuals with a single conviction, were not convicted throughout the observation period. This group accounts for an estimated 71% of the sampled population. A second group of individuals who ceased their offending (as measured by conviction) by their early 20s was labeled "adolescent limited" after Moffitt (1993). This group is estimated to constitute 22% of the population. Finally, a chronic group was identified and is estimated to make up 7% of the population. Individu-

---

<sup>1</sup> Predicted behavior is calculated as the expected value of the random variable depicting each group's behavior. Expected values are computed based on model coefficient estimates. For the application depicted in Figure 1, this expectation equals the antilog of Equation 1. For the application depicted in Figure 2, it is calculated according to the relationship provided in footnote 4. For the application depicted in Figure 3, it equals the binomial probability as computed by Equation 3. Actual behavior is computed as the mean behavior of all persons assigned to the various groups identified in estimation. As described hereinafter, the assignments are based on the posterior probability of group membership.

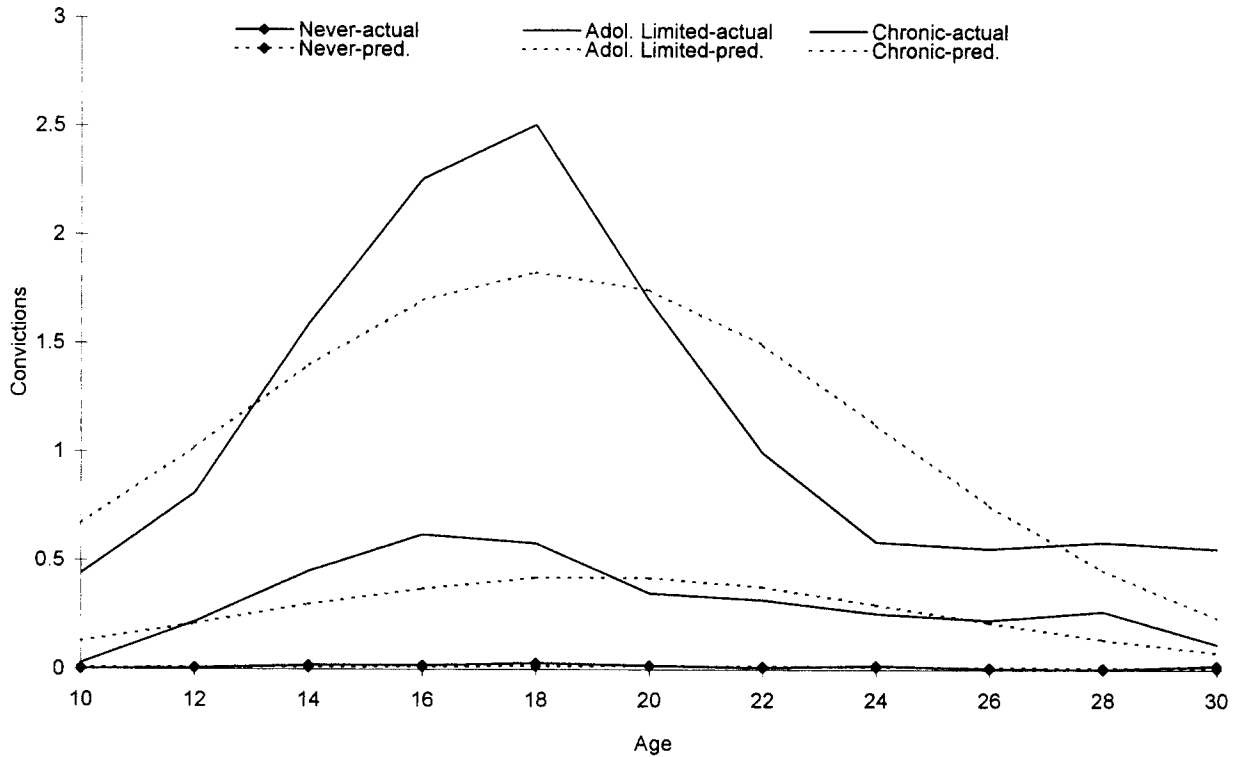


Figure 1. Trajectories of number of convictions (Cambridge sample). Adol. = adolescent; pred. = predicted.

als in this group offended at a high level throughout the observation period.

Figure 2 displays results based on the Montreal data. In this analysis the response variable was a psychometric scale of physical aggression. A four-group model was found to best fit the data. A group called "nevers" is composed of individuals who never display physically aggressive behavior to any substantial degree. This group is estimated to make up about 15% of the sample population. A second group that constitutes about 50% of the population is best labeled "low-level desisters." At age 6 boys in this group displayed modest levels of physical aggression, but by age 10 they had largely desisted. A third group, constituting about 30% of the population, was labeled "high-level near desisters." This group started off scoring high on physical aggression at age 6 but by age 15 scored far lower. Notwithstanding this marked decline, at age 15 they continued to display a modest level of physical aggression. Finally, there is a small chronic group, constituting about 5% of the population, who displayed high levels of physical aggression throughout the observation period.

Figure 3 displays another variant of the analysis of the physical aggression data. For this analysis the aggression data for each assessment period was transformed from an intensity scale over the interval 0 to 6 to a binary indicator equal to 1 if the individual displayed any evidence of physical aggression in that period (i.e., if his score was greater than 0). Such a symptom indicator variable is an example of binary data—a type of data that is widely analyzed in the social sciences. It is also the type of data Rindskopf's group-based method was designed to analyze. Three symptom trajectories are depicted in Figure 3. One group, constituting about 20% of the population, never displayed any symptoms. This group is the counterpart of the nevers in Figure 2. A second group includes about 50% of the population. It started off with a fairly high probability of displaying symptoms of physical aggression, about .6, but this probability declined rapidly to near 0 by age 15. The final group, which is estimated to account for about 30% of the population, began with a probability of physical aggression that is only modestly higher than that of the second group, but the subsequent rate of decline was

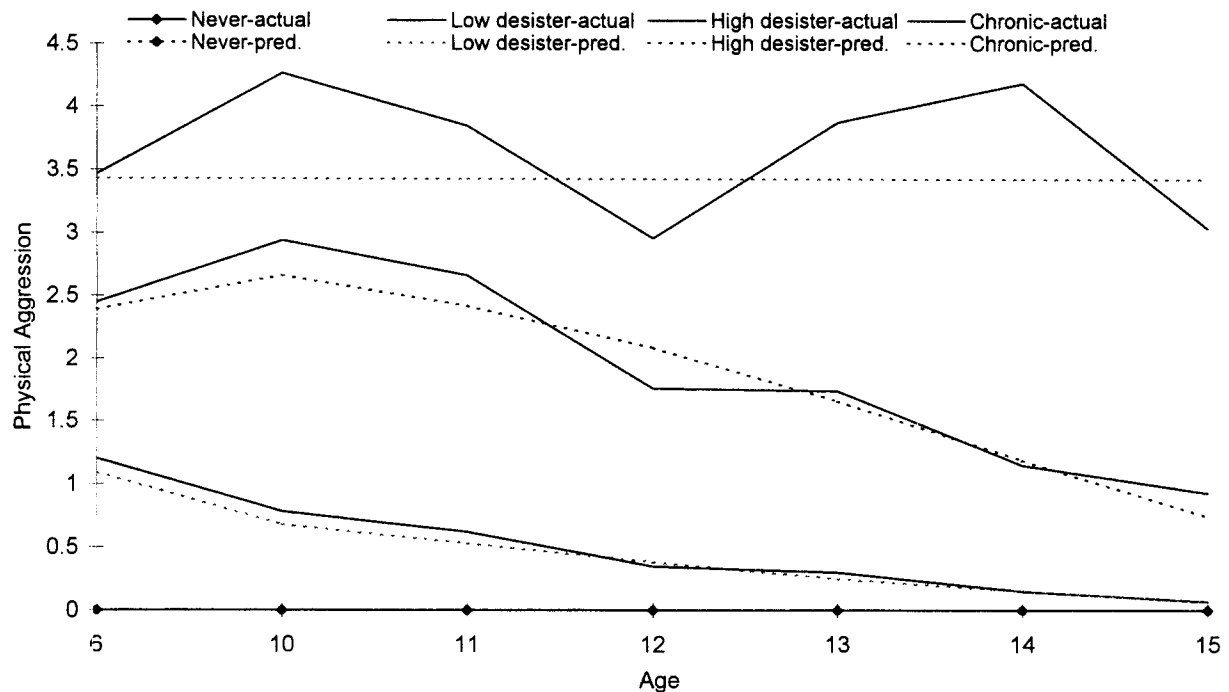


Figure 2. Trajectories of physical aggression (Montreal sample). pred. = predicted.

far slower. Even at age 15, the probability of their displaying symptoms of physical aggression is about 4.

### Model

Figure 4 provides an overview of the model and its key outputs. Model estimation does not require the *ex ante* sorting of the data among the various groups depicted in Figures 1, 2, and 3. To the contrary, the data are used to identify the number of groups that best fits the data and the shape of the trajectory for each group. The data also provide an estimate of the proportion of the population whose measured behavior conforms most closely to each trajectory group.

In the discussion that follows, I first lay out the general form of the model for any given number of groups. Next, I discuss the determination of the optimal number of groups and the shape of trajectory for each group. I then move to a discussion of how the model's parameter estimates can be used to compute the final element depicted in Figure 4—the probability that each individual  $i$  in the estimation sample belongs to each of the trajectory groups identified in estimation. These capabilities are illustrated with re-

sults from the Cambridge and Montreal data sets. For the Montreal data the censored normal analysis is used for this purpose. However, all of these capabilities are equally well employed in the analysis of binary data.

Differences in the form of the response variable in the analyses depicted in Figures 1, 2, and 3—a count variable, a psychometric scale, and a binary variable—necessitate technical differences in the statistical model used in each analysis. For the count data the underlying mixture model builds from the Poisson distribution and its even more general relative, the zero-inflated Poisson (Lambert, 1993). The Poisson family of distributions is widely used in the analysis of count data (Cameron & Trivedi, 1986; Parzen, 1960). For the psychometric scale data the underlying model is based on the censored normal distribution. The censored normal distribution is well suited to accommodate a common feature of psychometric scale data. Typically, a sizable contingent of the sample exhibits none of the behaviors measured by the scale. The result is a clustering of data at the scale minimum. Also, there is usually a smaller contingent that exhibits all of the behaviors measured by the scale. The result is another cluster of data at the scale

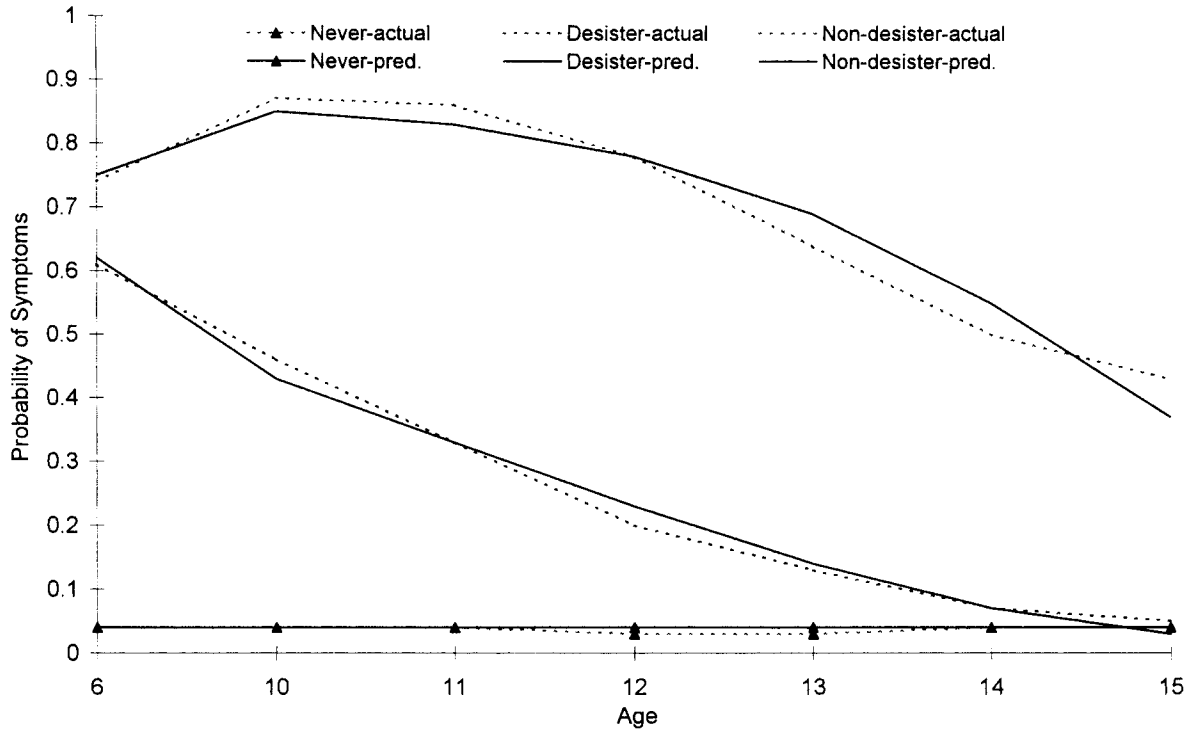


Figure 3. Trajectories of symptoms of physical aggression (Montreal sample). pred. = predicted.

maximum. Finally, the binary logit distribution is used to model binary data. See Maddala (1983) or Greene (1990) for a thorough discussion of the censored normal and binary logit distributions.

Like hierarchical and latent curve modeling, a polynomial relationship is used to model the link between age and behavior. Specifically, a quadratic relationship is assumed.<sup>2</sup> For the Poisson-based model it is assumed that

$$\log(\lambda_{it}^j) = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2, \quad (1)$$

where  $\lambda_{it}^j$  is the expected number of occurrences of the event of interest (e.g., convictions) of subject  $i$  at time

$t$  given membership in group  $j$ .  $\text{Age}_{it}$  is subject  $i$ 's age at time  $t$ , and  $\text{Age}_{it}^2$  is the square of subject  $i$ 's age at time  $t$ .<sup>3</sup> The model's coefficients— $\beta_0^j$ ,  $\beta_1^j$ , and  $\beta_2^j$ —determine the shape of the trajectory and are superscripted by  $j$  to denote that the coefficients are not constrained to be the same across the  $j$  groups. The conditional probability of the actual number of events,  $P(y_{it}^j|j)$ , given  $j$  is assumed to follow the well-known Poisson distribution.<sup>4</sup>

For the censored normal model, the linkage between age and behavior is established by means of a latent variable,  $y_{it}^j$ , that can be thought of as measur-

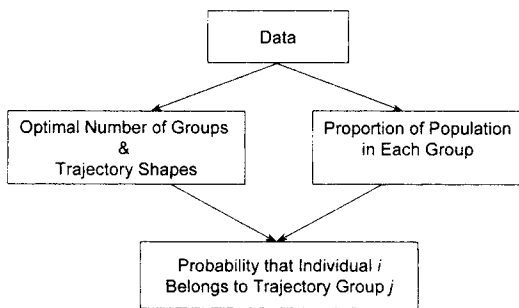


Figure 4. Overview of the model.

<sup>2</sup> The software package used in estimating the models reported here allows for estimation of up to a cubic polynomial in age. For ease of exposition I limit the discussion to examples where the maximum order considered is quadratic.

<sup>3</sup> A log-linear relationship between  $\lambda_{it}^j$  and age is assumed to ensure that the requirement that  $\lambda_{it}^j > 0$  is fulfilled in model estimation.

<sup>4</sup> In an even more general version of this model,  $P(y_{it}^j|j)$  is assumed to follow the zero-inflated Poisson distribution. See Land, McCall, & Nagin (1996), Nagin and Land (1993), or Roeder, Lynch, and Nagin (in press) for the development of this more general case.

ing potential for engaging in the behavior of interest—say, physical aggression—for individual  $i$ 's age at time  $t$  given membership in group  $j$ . Again, a quadratic relationship is assumed between  $y_{it}^{*j}$  and age:

$$y_{it}^{*j} = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2 + \varepsilon_{it}, \quad (2)$$

where  $\text{Age}_{it}$  and  $\text{Age}_{it}^2$  are as previously defined and  $\varepsilon$  is a disturbance assumed to be normally distributed with zero mean and constant variance  $\sigma^2$ .

The latent variable,  $y_{it}^{*j}$ , is linked to its observed but censored counterpart,  $y_{it}^j$ , as follows. Let  $S_{\min}$  and  $S_{\max}$ , respectively, denote the minimum and maximum possible score on the measurement scale. The model assumes

$$\begin{aligned} y_{it} &= S_{\min} \text{ if } y_{it}^{*j} < S_{\min}, \\ y_{it} &= y_{it}^{*j} \text{ if } S_{\min} \leq y_{it}^{*j} \leq S_{\max}, \text{ and} \\ y_{it} &= S_{\max} \text{ if } y_{it}^{*j} > S_{\max}. \end{aligned}$$

In words, if the latent variable,  $y_{it}^{*j}$ , is less than  $S_{\min}$ , it is assumed that observed behavior equals this minimum. Likewise, if the latent variable,  $y_{it}^{*j}$ , is greater than  $S_{\max}$ , it is assumed that observed behavior equals this maximum. Only if  $y_{it}^{*j}$  is within the scale minimum and maximum does  $y_{it} = y_{it}^{*j}$ .<sup>5,6</sup>

Finally, consider the case where the measured response,  $y_{it}$ , is binary. In this case it is assumed that conditional on membership in group  $j$  the probability that  $y_{it} = 1$ ,  $\alpha_{it}^j$ , follows the binary logit distribution:

$$\alpha_{it}^j = \frac{e^{\beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2}}{1 + e^{\beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2}}. \quad (3)$$

In the Poisson, censored normal, and binary logit formulations, the parameters defining the shape of the trajectory— $\beta_0^j$ ,  $\beta_1^j$ , and  $\beta_2^j$ —are left free to differ across groups. This flexibility is a key feature of the model because it allows for easy identification of population heterogeneity not only in the level of behavior at a given age but also in its development over time. Figure 5 illustrates two hypothetical possibilities. A single peaked trajectory—Trajectory A—is implied if  $\beta_1 > 0$  and  $\beta_2 < 0$ . Thus, if data collection began at age 1, the trajectory would imply that for this group the occurrence of the behavior rose steadily until age 6 and then began a steady decline. Alternatively, if data collection began at age 6, as was the case in the Montreal study, generally it would be inappropriate to extrapolate backward to a younger age outside the period of measurement. Thus, for a model

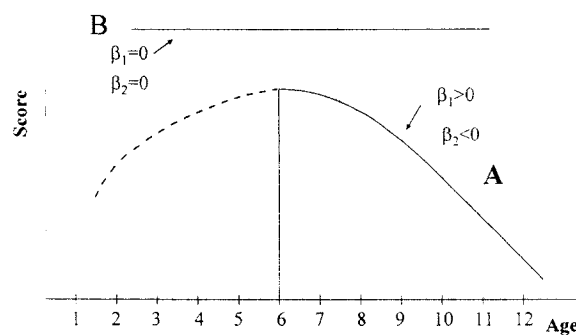


Figure 5. Two hypothetical trajectories. A = single peaked; B = chronic trajectory.

based on data from age 6 onward, this trajectory would imply a steady decline in the behavior following the initial assessment. Such a trajectory would typify desistance from the behavior. The second trajectory (B) depicted in Figure 5 has no curvature. Rather it remains constant over age. This trajectory is implied if  $\beta_1 = 0$  and  $\beta_2 = 0$ . If that stable level of the behavior is high, this trajectory would typify a group that chronically engages in the behavior. Other interesting possibilities include trajectories in which growth is either steadily accelerating or decelerating. The former would be characterized by a trajectory in which both  $\beta_1$  and  $\beta_2$  are positive and the latter by both being negative.

The trajectories depicted in Figures 1, 2, and 3 are the product of maximum-likelihood estimation. The likelihood function was constructed as follows. Let the vector  $\mathbf{Y}_i = \{y_{i1}, y_{i2}, \dots, y_{iT}\}$  denote the longitudinal sequence of individual  $i$ 's behavioral measure-

<sup>5</sup> The expected value of the latent variable  $y_{it}^{*j}$  equals  $\beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2$ , which is denoted below by  $\beta x_j$ . The expected value of the measured quantity,  $E(y_{it}^j)$ , assuming  $j$  was observed, is  $E(y_{it}^j) = \Phi_{\min}^j S_{\min} + \beta' x_j (\Phi_{\max}^j - \Phi_{\min}^j) + \sigma (\phi_{\min}^j - \phi_{\max}^j) + (1 - \Phi_{\max}^j) S_{\max}$ , where  $\Phi^{(*)}$  denotes the cumulative normal distribution function and  $\Phi_{\min}$  and  $\Phi_{\max}$ , respectively, denote  $\Phi^j[(S_{\min} - \beta x_j^j)/\sigma]$  and  $\Phi^j[(S_{\max} - \beta x_j^j)/\sigma]$ , with corresponding definitions for  $\phi_{\min}$  and  $\phi_{\max}$ , where  $\phi$  denotes the normal density function. This relationship was used to compute the predicted components of Figure 1.

<sup>6</sup> I use the term *latent variable* to describe  $y_{it}^{*j}$  because it is not fully observed. Thus, my use of the term *latent* is different from that in the psychometric literature, where the term *latent factor* refers to an unobservable construct that is assumed to give rise to multiple manifest variables.

ments over the  $T$  periods of measurement,  $P^j(Y_i)$  denote the probability of  $Y_i$  given membership in group  $j$ , and  $\pi_j$  denote the probability of membership in group  $j$ . For count data  $P^j(Y_i)$  is constructed from the Poisson distribution, for psychometric scale data from the censored normal distribution, and for the binary data from the binary logit distribution. If group membership was observed, the sampled individuals could be sorted by group membership and their trajectory parameters estimated with readily available Poisson, censored normal (tobit), and logit regression software routines (e.g., STATA, 1995 or LIMDEP, Greene, 1991).<sup>7</sup>

However, group membership is not observed. Indeed the proportion of the population composing group  $j$ ,  $\pi_j$ , is an important parameter of interest in its own right. Thus, construction of the likelihood requires the aggregation of the  $J$  conditional likelihoods,  $P^j(Y_i)$ , to form the unconditional probability of the data,  $P(Y_i)$ :

$$P(Y_i) = \sum_j \pi_j P^j(Y_i), \quad (4)$$

where  $P(Y_i)$  is the unconditional probability of observing individual  $i$ 's longitudinal sequence of behavioral measurements. It equals the sum across the  $J$  groups of the probability of  $Y_i$  given membership in group  $j$  weighted by the proportion of the population in group  $j$ . The log of the likelihood for the entire sample is thus the sum across all individuals that compose the sample of the log of Equation 4 evaluated for each individual  $i$ . The parameters of interest— $\beta_0^j$ ,  $\beta_1^j$ ,  $\beta_2^j$ ,  $\pi_j$ , and, in addition,  $\sigma$  for the censored normal—can be estimated by maximization of this log likelihood. As previously mentioned, SAS-based software for accomplishing this task is available on request.

For a derivation of the likelihood see the appendix, but intuitively, the estimation procedure works as follows. Suppose unbeknownst to us there were two distinct groups in the population: youth offenders, constituting 50% of the population, who up to age 18 have an expected offending rate,  $\lambda$ , of 5 and who after age 18 have a  $\lambda$  of 1, and adult offenders, constituting the other 50% of the population, whose offending trajectory is the reverse of that of the youth offenders—through age 18 their  $\lambda = 1$ , and after age 18 their  $\lambda$  increases to 5. If we had longitudinal data on the recorded offenses of a sample of individuals from this population, we would observe two distinct groups: a clustering of about 50% of the sample who generally have many offenses prior to age 18 and

relatively few offenses after age 18 and another 50% clustering with just the reverse pattern.

Suppose these data were analyzed under the assumption that the relationship between age and  $\lambda$  was identical across all individuals. The estimated value of  $\lambda$  would be a "compromise" estimate of about 3 for all ages from which we would mistakenly conclude that in this population the rate of offending is invariant with age. If the data were analyzed using the approach described here, which specifies the likelihood function as a mixing distribution, no such mathematical compromise would be necessary. The parameters of one component of the mixture would effectively be used to accommodate (i.e., match) the youth offending portion of the data whose offending declines with age, and another component of the mixing distribution would be available to accommodate the adult offender data whose offending increases with age.

## Model Selection

This section addresses two important issues in model selection: (a) determination of the optimal number of groups to compose the mixture and (b) determination of the appropriate order of the polynomial used to model each group's trajectory. Here *order* refers to the degree of the polynomial used to model the group's trajectory, where a second-order trajectory is defined by a quadratic equation, a first-order trajectory is defined by a linear equation in which  $\beta_2$  is set equal 0, and a zero-order trajectory is defined by a flat line in which  $\beta_1$  and  $\beta_2$  are set equal to zero. I discuss these two issues in turn.

One possible choice for testing the optimality of a specified number of groups is the likelihood ratio test. However, the null hypothesis (i.e., three components vs. more than three components) is on the boundary of the parameter space, and hence the classical asymptotic results that underlie the likelihood ratio test do not hold (Erdfelder, 1990; Ghosh & Sen, 1985; Titterington, Smith, & Makov, 1985). The likelihood ratio test is suitable only for model selection problems in which the alternative models are nested. In mixture models, a  $k$  group model is not nested within a  $k + 1$

<sup>7</sup> In the econometric literature censored normal regression is generally called *tobit regression* after its originator, James Tobin (Tobin, 1958).



group model, and, therefore, it is not appropriate to use the likelihood ratio test for model selection.<sup>8</sup>

Given these problems with the use of the likelihood ratio test for model selection, we follow the lead of D'Unger, Land, McCall, and Nagin (1998) and use the Bayesian information criterion (BIC) as a basis for selecting the optimal model. For a given model, BIC is calculated as follows:

$$\text{BIC} = \log(L) - 0.5 * \log(n) * (k), \quad (5)$$

where  $L$  is the value of the model's maximized likelihood,  $n$  is the sample size, and  $k$  is the number of parameters in the model. Kass and Raftery (1995) and Raftery (1995) have argued that BIC can be used for comparison of both nested and unnested models under fairly general circumstances. When prior information on the correct model is limited, they recommended selection of the model with the maximum BIC. Note that BIC is always negative, so the maximum BIC will be the least negative value. In even more recent work, Keribin (1997) demonstrated that BIC identifies the optimal number of groups in finite mixture models. Her result is specifically relevant for the mixture models demonstrated here.

For insight into the usefulness of BIC as a criterion for model selection, consider its calculation. The first term of Equation 5 is always negative. For a model that predicts the data perfectly it equals zero. As the quality of the model's fit with the data declines, this term declines (i.e., becomes more negative). One way to improve fit and thereby reduce the first term is to add more parameters to the model. The second term extracts a penalty proportional to the log of the sample size for the addition of more parameters. Thus, on the basis of the BIC criterion, expansion of the model by addition of a trajectory group is desirable only if the resulting improvement in the log likelihood exceeds the penalty for more parameters. As Kass and Raftery (1995) noted, the BIC rewards parsimony. For this application the BIC criterion will tend to favor models with fewer groups.

Table 1 shows BIC scores for models with varying numbers of groups. For the Cambridge data the pattern is seemingly very distinct—BIC appears to reach a clear maximum at three groups. I say "seemingly," because, without a concrete standard for calibrating the magnitude of the change in BIC, it is difficult to calibrate what constitutes a distinct maximum.

Kass and Wasserman (1995) and Schwarz (1978) have provided such a standard. Let  $B_{ij}$  denote the Bayes factor comparing models  $i$  and  $j$ , where for

Table 1  
*BIC-Based Calculations of the Probability That a "j" Group Model Is the Correct Model for Different Numbers of Groups of Quadratic Trajectories*

No. of groups	Data set			
	Cambridge		Montreal	
	BIC	Probability correct model	BIC	Probability correct model
2	-1583.43	.00	-7325.26	.00
3	-1552.62	1.00	-7289.52	.43
4	-1569.42	.00	-7289.27	.55
5	-1586.21	.00	-7292.54	.02

Note. BIC = Bayesian information criterion.

this application model  $i$  might be a two-group model and  $j$  a three-group model. The Bayes factor measures the odds of each of the two competing models being the correct model. It is computed as the ratio of the probability of  $i$  being the correct model to  $j$  being the correct model. Thus, a Bayes factor of 1 implies that the models are equally likely, whereas a Bayes factor of 10 implies that model  $i$  is 10 times more likely than  $j$ . Table 2 shows Jeffreys's scale of evidence for Bayes factors as reported by Wasserman (1997).<sup>9</sup>

Computation of the Bayes factor is in general very difficult and indeed commonly impossible. Schwarz (1978) and Kass and Wasserman (1995), however, have shown that  $e^{\text{BIC}_i - \text{BIC}_j}$  is a good approximation of the Bayes factor for problems in which equal weight is placed on the prior probabilities of models  $i$  and  $j$ . On the basis of this approximation, for the Cambridge

<sup>8</sup> The problem is most easily illustrated with an example. The likelihood ratio test is computed as minus two times the difference in the likelihood of the two nested models. This statistic is asymptotically distributed as chi-squared with degrees of freedom equal to the difference in number of parameters between the nested models. For our mixture model, the degrees of freedom is indeterminate, because a group can become superfluous in two ways. One is by the proportion of the population in that group,  $\pi_j$ , approaching zero. Alternatively, the three parameters defining the trajectory of one group can collapse onto those for another group. What then is the appropriate degrees of freedom—one or three?

<sup>9</sup> Jeffreys was an early and very prominent contributor to Bayesian statistics.

Table 2  
*Jeffreys's Scale of Evidence for Bayes Factors*

Bayes factor	Interpretation
$B_{ij} < 1/10$	Strong evidence for model $j$
$1/10 < B_{ij} < 1/3$	Moderate evidence for model $j$
$1/3 < B_{ij} < 1$	Weak evidence for model $j$
$1 < B_{ij} < 3$	Weak evidence for model $i$
$3 < B_{ij} < 10$	Moderate evidence for model $i$
$B_{ij} > 10$	Strong evidence for model $i$

*Note.* Adapted from "Bayesian Model Selection and Model Averaging" by L. Wasserman, 1997, Working Paper No. 666, Carnegie Mellon University, Department of Statistics. Copyright 1997 by L. Wasserman. Adapted with permission.

data the odds of the three-group model compared with the two- or four-group model far exceed 1000 to 1. Thus, according to Jeffreys's scale this is very strong evidence in favor of the three-group model.<sup>10</sup>

Schwarz (1978) and Kass and Wasserman (1995) have also provided a related metric for comparing more than two models. Let  $p_j$  denote the posterior probability that model  $j$  is the correct model, where in general  $j$  is greater than 2. They show that  $p_j$  is reasonably approximated by the following:

$$p_j = \frac{e^{\text{BIC}_j - \text{BIC}_{\max}}}{\sum_j e^{\text{BIC}_j - \text{BIC}_{\max}}}, \quad (6)$$

where  $\text{BIC}_{\max}$  is the maximum BIC score of the models under consideration.

Also shown in Table 1 are the probabilities, as computed using Equation 6, that the models with varying numbers of groups are the true model. With this BIC-based probability approximation, for the Cambridge data the probability of the three-group model is near 1.

Consider now the BIC scores for the models estimated with the Montreal data. The four-group model has the best BIC score, but the probability of its being the correct model, .55, is far less than 1. Although the four-group model is far more likely than the five-group model, the three-group model is a close competitor. Its probability is .43. On the basis of Jeffreys's scale, the four-group model is strongly preferred to the five-group model. The odds ratio in favor of the four-group model is 26 to 1 (i.e., .55/.02). However, the edge compared with the three-group model is slight with an odds ratio of only 1.28 (i.e., .55/.43). Thus, for models in which each trajectory is described by a full quadratic specification, the three- and four-group models fit the data about equally well.

The Montreal analysis illustrates a situation in which the mechanical application of the BIC model selection criterion does not result in an unambiguous determination of the "best" model. This brings me to the issue of determination of the appropriate order for modeling the trajectory of each group that makes up the mixture. The best model will not necessarily involve a mixture in which all trajectories are of the same order (e.g., quadratic). In principle one could exhaustively explore all possible combinations of orders for a model. In practice this approach is generally not practical. For a four-group model alone, there are 81 (i.e.,  $3^4$ ) possible combinations of trajectories models, the number of possibilities in a four-group model is 256. Further, a full search would require estimating models over varying numbers of groups. To reduce the number of alternatives, an analyst will generally have to use knowledge of the problem domain to limit the model search process. For instance, in the Montreal example substantive considerations suggest that an improved, more parsimonious model should restrict the number of parameters used to describe the never and chronic trajectories. Description of a never trajectory does not require a three-parameter model: a zero-order model with a "very negative" intercept should suffice. Indeed the large standard errors (not reported) associated with the parameters of the never trajectory strongly suggest that the trajectory is overparameterized.

Table 3 shows the probabilities that various forms of three- and four-group models are the correct model. We refer to models in which the trajectories for all groups are quadratic as *Type A* models. Three- and four-group models in which the never trajectory is described by a zero-order model are labeled *Type B* models. Also shown are three- and four-group models in which, in addition, the chronic trajectory is described by a zero-order model (*Type C*). The information was prompted by the theories of Moffitt (1993) and Patterson et al. (1989), which predict the existence of a small group of chronically antisocial individuals. The BIC-based probability calculations provide strong support for the four-group, Type C model in Table 3. Compared with the other options

<sup>10</sup> Using the BIC criterion, D'Unger et al. (1998) and Roeder, Lynch, and Nagin (in press) showed that a four-group model is best. In those analyses the model is fit based on the more general zero-inflated Poisson.

Table 3  
*BIC-Based Calculation of Probability of Correct Model for Models Combining Quadratic and Nonquadratic Trajectories: Montreal Data*

No. of groups	Model type		
	A	B	C
3	.00	.00	.00
4	.00	.07	.93

*Note.* BIC = Bayesian information criterion; Model Type A = all trajectories quadratic; Model Type B = one single parameter trajectory, remaining quadratic; Model Type C = two single parameter trajectories, remaining quadratic.

listed in the table, the probability of this being the correct model is .93. The closest competitor is the four-group, Type B model. However, the posterior probability of it being the correct model is only .07.

As the Montreal analysis illustrates, the determination of the optimal number of groups is not always a clear-cut process. Although Bayesian statisticians have made important strides in developing theoretically grounded criteria for model selection, the results are recent and not yet widely used. Further, as illustrated by the Montreal example, the model search process may also be guided in part by nonstatistical considerations. This injects a degree of subjectivity into model selection. Still, it is important to recognize that use of the BIC criterion for model selection adds a very significant degree of statistical objectivity to the determination of the optimal number of groups. This provides an important check against the spurious interpretation of random fluctuations in the data as reflecting systematic patterns of behavior.

#### Calculation and Use of Posterior Group Membership Probabilities

It is not possible to determine definitively an individual's group membership. However, it is possible to calculate the probability of his or her membership in the various groups that make up the model. These probabilities, the posterior probabilities of group membership, are among the most useful products of the group-based modeling approach. Specifically, based on the model coefficient estimates, for each individual  $i$  the probability of membership in group  $j$  is calculated on the basis of the individual's longitudinal pattern of behavior,  $Y_i$ . We denote this probability by  $\hat{P}(j|Y_i)$ . It is computed as follows:

$$\hat{P}(j|Y_i) = \frac{\hat{P}(Y_i|j)\hat{\pi}_j}{\sum_j \hat{P}(Y_i|j)\hat{\pi}_j}, \quad (7)$$

where  $\hat{P}(Y_i|j)$  is the estimated probability of observing  $i$ 's actual behavioral trajectory,  $Y_i$ , given membership in  $j$ , and  $\hat{\pi}_j$  is the estimated proportion of the population in group  $j$ . The quantity  $\hat{P}(Y_i|j)$  can be calculated postestimation based on the maximum-likelihood estimates of the trajectory parameters.

The posterior probability calculations provide the researcher with an objective basis for assigning individuals to the development trajectory group that best matches their behavior. Individuals can be assigned to the group to which their posterior membership probability is largest. On the basis of this maximum posterior probability assignment rule, an individual from the Cambridge sample with a cluster of offenses in adolescence but none thereafter, in all likelihood, will be assigned to the adolescent limited category. Alternatively, an individual who offends at a comparatively high rate throughout the observation period will likely be assigned to the chronic group.

Table 4 shows the mean assignment probability for the Montreal and Cambridge data. For example, in the Cambridge data the mean chronic group posterior probability for the 27 individuals assigned to this group was very high—.95. The counterpart average for the 77 individuals assigned to the adolescent limited group is similarly large—.94. For the Montreal data, classification certainly is not so high but still seems reasonably good. Across the four groups it ranges from .73 to .93.

One informative use of the posterior probability-based classifications is to create profiles of the "average" individual following the trajectory characterized by each group. Table 5 shows summary statistics on individual characteristics and behaviors of each group for both data sets. The profiles conform with longstanding findings on risk and protective factors for antisocial and problem behaviors. In the Cambridge data those in the chronic group, on average, were most likely to have lived in a low-income household, to have had at least one parent with a criminal record, to have been subjected to poor parenting, and to have displayed a high propensity to engage in risky activities. Conversely, the never group was lowest on these risk factors. The pattern is similar for physical aggression. Members of the chronic group had the least well-educated parents and most frequently scored in the lowest quartile of the measured IQ distribution of the sample, whereas the nevers were high-

Table 4  
Average Assignment Probability Conditional on Assignment by Maximum Probability Rule

Assigned group	Group			
	Never	Adolescent limited	Chronic	
Cambridge data				
Never	.94	.06	.00	
Adolescent limited	.03	.94	.03	
Chronic	0	.05	.95	
Montreal data				
	Never	Low desister	High desister	Chronic
Never	.73	.27	.00	.00
Low desister	.00	.93	.10	.00
High desister	.00	.09	.84	.06
Chronic	.00	.00	.24	.76

est on these protective factors. Further, 90% of those in the chronic group failed to reach the eighth grade on schedule, and 13% had a juvenile record by age 18; only 19% of the nevers had fallen behind by the eighth grade, and none had a juvenile record. In between are the low-level and high-level desisters, who themselves order as expected on these characteristics and behaviors.

Aside from providing a basis for group assignment, the posterior probabilities can provide an objective criterion for selecting subsamples for follow-up data collection. For example, in an analysis based on the Montreal data, Nagin and Tremblay (in press) found that individuals following the chronic physical aggression trajectory depicted in Figure 3 displayed

heightened levels of violence at age 18, controlling for chronic opposition and hyperactivity. Conversely, it was found that chronic opposition and hyperactivity did not predict violence at age 18, controlling for chronic physical aggression. Nagin and Tremblay concluded that chronic physical aggression was a distinct risk factor for later violence. A follow-up study that is currently under way at the time of this writing was devised to examine whether self-regulatory processes were a possible explanation for the findings. The study involves a series of laboratory assessments of selected individuals in the Montreal study. Individuals were recruited for this study based on their posterior probabilities of membership in the various trajectory groups for physical aggression, opposition,

Table 5  
Group Profile

Variable	Group			
	Never	Adolescent limited	Chronic	
Cambridge data				
Low household income (%)	16.4	35.1	55.6	
Poor parenting (%)	17.4	33.8	51.9	
High risk taking (%)	21.4	42.9	74.1	
Parents with criminal record (%)	17.4	42.9	63.0	
Montreal data				
	Never	Low desister	High desister	Chronic
Years of school—mother	11.1	10.8	9.8	8.4
Years of school—father	11.5	10.7	9.8	9.1
Low IQ <sup>a</sup> (%)	21.6	26.8	44.5	46.4
Completed eighth grade on time (%)	80.3	64.6	31.8	6.5
Juvenile record (%)	0.0	2.0	6.0	13.3
No. of sexual partners at age 17 (past year)	1.2	1.7	2.2	3.5

<sup>a</sup> Measured IQ score is in the lowest quartile of the sample.

and hyperactivity. The purpose was to recruit strategically to ensure that the sample included persons with distinct combinations of development in externalizing behaviors (e.g., high probability of chronic physical aggression but low probability of chronic opposition). The posterior probabilities provide objective, quantified criteria for such strategic sampling.

### Statistically Linking Group Membership to Covariates

What individual, familial, and environmental factors distinguish the populations of the various trajectory groups? Are the factors consistent with received theory on developmental trajectories? What statistical procedures are most appropriate to test whether such factors distinguish among the trajectory groups? Questions such as these follow naturally from the identification of distinct groups of developmental trajectories.

The profiles reported in Table 5 are a first step in addressing these questions but indeed only a beginning. First, the profiles are simply a collection of univariate contrasts. For the purpose of constructing a more parsimonious list of predictors or for causal inference, a multivariate procedure is required to sort out redundant predictors and to control for potential confounders. Second, the profiles are based on group identifications that are probabilistic, not certain. Conventional statistical methods to test for cross-group differences, such as *F*- and chi-square-based tests, assume no classification error in group identification (Roeder et al., in press). Thus, in general they are technically inappropriate in this setting.<sup>11</sup>

The mixture model of developmental trajectories comprises two basic components: (a) an expected trajectory given membership in group *j* and (b) a probability of group membership denoted by  $\pi_j$ . Thus far, the discussion has focused on the former component. The discussion of the test for factors that distinguish groups turns our focus to the latter component. The profiles in Table 5 show that the trajectory groups are composed of individuals who differ in substantial and predictable ways. Provision of the capacity to test formally for whether and by what degree such factors distinguish the groups requires that the model be generalized to allow  $\pi_j$  to depend on characteristics of the individual. Heretofore,  $\pi_j$  has been described as the proportion of the population following each trajectory group *j*. Equivalently, it can be described as the probability that a randomly chosen individual from the

population under study belongs to trajectory group *j*. By allowing  $\pi_j$  to vary with individual characteristics, it is possible to test whether and by how much a specified factor affects probability of group membership controlling for the level of other factors that potentially affect  $\pi_j$ .

Let  $x_i$  denote a vector of factors measuring individual, familial, or environmental factors that potentially are associated with group membership, and let  $\pi_j(x_i)$  denote the probability of membership in group *j* given  $x_i$ . For a two-group model the logit model is a natural candidate for modeling group membership probability as a function of  $x_i$ . For this special case, we need only estimate  $\pi_j(x_i)$  for one group, say Group 1, because  $\pi(x_i) = 1 - \pi_1(x_i)$ , where

$$\pi_1(x_i) = \frac{e^{x_i\theta}}{1 + e^{x_i\theta}}.$$

For the more general case, in which there are more than two groups, the logit model generalizes to the multinomial logit model (Maddala, 1983):

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_j e^{x_i\theta_j}}, \quad (8)$$

where the parameters of the multinomial logit model,  $\theta_j$ , capture the impact of the covariates of interest,  $x_i$ , on probability of group membership. Without loss of generality,  $\theta_j$  for one "contrast" group can be set equal to zero. The coefficient estimates for the remaining groups should be interpreted as measuring the impact of covariates on group membership relative to the contrast group.

Table 6 illustrates the application of the extended model to the Montreal data. Results are reported for a four-group model in which probability of group membership is related to four variables: one parent having less than a ninth-grade education, both parents having less than a ninth-grade education, having a mother who began childbearing as a teenager, and scoring in the lowest quartile of the measured IQ of the sampled boys. The first panel of Table 6 shows coefficient estimates and *t* statistics. For this example, the never-convicted group serves as the contrast group. Con-

<sup>11</sup> Roeder et al. (in press) explored the impact of classification error on statistical inference. One finding is not surprising. When classification is highly certain, as is the case in the Cambridge data, errors in inference using conventional methods are small.

Table 6  
*The Impact of Parental Education, Low IQ, and Teen Onset of Motherhood on Group Membership Probabilities:  
 Montreal Data*

Variable-condition	Group				
	Never	Low desister	High desister	Chronic	
Multinomial logit coefficients (with $\pm$ statistics given in parentheses)					
Constant	—	1.04 (6.36)	-0.02 (-0.10)	-2.14 (-4.49)	
Low education <sup>a</sup> —parent	—	0.44 (1.46)	0.69 (2.28)	-0.19 (-0.30)	
Low education <sup>a</sup> —both parents	—	0.18 (0.48)	0.67 (1.83)	0.98 (1.73)	
Low IQ <sup>b</sup>	—	0.07 (0.22)	0.85 (2.85)	0.92 (1.83)	
Teen mom <sup>c</sup>	—	0.89 (2.03)	1.10 (2.57)	2.27 (3.89)	
Predicted membership probabilities based on multinomial logit model coefficient estimates					
No risk factors		.20	.57	.20	.02
Low education <sup>a</sup> —both parents—only		.15	.51	.29	.05
Low IQ <sup>b</sup> only		.15	.46	.35	.04
Teen mom <sup>c</sup> only		.08	.58	.25	.09
Low education <sup>a</sup> —both parents—and low IQ <sup>b</sup> and teen mom <sup>c</sup>		.03	.29	.44	.25

<sup>a</sup> Parent has less than a ninth-grade education.

<sup>b</sup> Measured IQ score is in the lowest quartile of the sample.

<sup>c</sup> Participant's mother began childbearing as a teenager.

sider first the parental low-education variables. Relative to the never group, having one or more poorly educated parents does not significantly increase the probability of membership in the low-level desister group ( $\alpha = .05$ , one-tailed test). However, probability of membership in the high-level near-desister and chronic groups is significantly increased if both parents are poorly educated. Low IQ has a similar impact on group membership probability. Compared with the never group, probability of membership in the low-level group is not significantly affected, whereas probability of membership in the two higher aggression trajectories is significantly increased. Finally, the teen mom variable is associated with a significant increase in all group membership probabilities relative to the never group.

The second panel of Table 6 shows calculations of the predicted probabilities of group membership based on the coefficient estimates. The calculations were performed by substituting the coefficient estimates into Equation 8 and then computing group membership probabilities for assumed values of  $x_i$ . The calculations show that if none of the risk factors are present, the probability of the never or low-level groups is .77 and the probability of the chronic group is only .02. The largest single factor affecting the chronic group membership probability is the teen mom variable. For the case in which an individual has none of the other risk factors except for a mother who

began childbearing as a teenager, the combined probability of the never and the low-level desister group declines to .66 and the probability of the chronic group increases to .09. Also reported is a high-risk scenario in which three risk factors are present—both parents poorly educated, low IQ, and teen motherhood. In this case the predicted probability of the chronic group is .25, an order of magnitude larger than the no-risk-factor case. Also, the predicted probability of the high-level near-desister group of .44 is double the probability for the no-risk-factor case.

Simultaneous estimation of the group-specific trajectory parameters and the group membership probabilities conditional on individual factors is easily accomplished with the above-referenced software. The efficiency of estimation is also relatively good. Although estimation time will depend on data set size, number of groups and covariates, and the speed of the computer, computation time is generally 5 to 10 min. For analyses in which hypotheses are well formed, such computation time is inconsequential. However, in analyses in which hypotheses are less well formed or analyses that are generally exploratory, much more effort will be put into model fitting. In these circumstances computation times of 5 to 10 min per run may be very cumbersome.

Two alternatives for highly efficient exploratory model fitting are recommended. Both are two-stage approaches with the same first stage—identification

of the best fitting model in terms of number of groups. The first stage involves a search for the best fitting model. This search is conducted without covariates distinguishing group membership. One alternative for the second-stage analysis makes use of the group membership assignments based on the maximum posterior probability rule. Multinomial logit models are then estimated relating group assignment to individual-level factors. A second alternative for the second-stage analysis is to regress group membership probabilities on the individual-level factors of interest. Both of these approaches for identifying candidate factors that distinguish group membership are highly efficient and easily conducted with conventional statistical software. Personal experience has shown that the results of an exploratory analysis conducted by either of these methods provide a good guide for specification of the proper model in which trajectories and the impact of covariates on membership probabilities are jointly estimated.<sup>12</sup>

### Comparative Advantages and Disadvantages of Group-Based Modeling

In this article I have described a group-based modeling strategy for analyzing developmental trajectories. Alternative methods generally treat the population distribution of development as continuous. The two leading examples of this continuous modeling strategy are hierarchical modeling and latent curve modeling.<sup>13</sup> The group-based and continuous modes of analysis each have their distinctive strengths. Growth curve modeling, whether in the hierarchical or latent variable tradition, is designed to identify average developmental tendencies, to calibrate variability about the average, and to explain that variability in terms of covariates of interest. By contrast, the mixture modeling strategy is designed to identify distinctive, prototypical developmental trajectories within the population, to calibrate the probability of population members following each such trajectory, and to relate those probabilities to covariates of interest.

Raudenbush (in press) offered a valuable perspective on the types of problems for which these two broad classes of modeling strategies are most suitable. He observed, "In many studies it is reasonable to assume that all participants are growing according to some common function but that the growth parameters vary in magnitude" (p. 30). He offered children's vocabulary growth curves as an example of such a growth process. Two distinctive features of such de-

velopmental processes are (a) they are generally monotonic—thus, the term *growth*—and (b) they vary regularly within the population. For such processes it is natural to ask, "What is the typical pattern of growth within the population and how does this typical growth pattern vary across population members?" Hierarchical and latent curve modeling are specifically designed to answer such a question.

Raudenbush (in press) also offered an example of a developmental process—namely, depression—that does not generally change monotonically over time and does not vary regularly through the population. He observed, "It makes no sense to assume that everyone is increasing (or decreasing) in depression. . . . many persons will never be high in depression, others will always be high, while others will become increasingly depressed" (p. 30). For problems such as this, he recommended the use of a multinomial-type method such as that demonstrated here. The reason for this recommendation is that the development does not vary regularly across population members. Instead trajectories vary greatly across population subgroups both in terms of the level of behavior at the outset of the measurement period and in the rate of growth and decline over time. For such problems a modeling strategy designed to identify averages and explain variability about that average is far less useful than a group-based strategy designed to identify distinctive clusters of trajectories and to calibrate how characteristics of individuals and their circumstances affect membership in these clusters.

### Summary and Discussion

This article has demonstrated a semiparametric, group-based approach for analyzing developmental

---

<sup>12</sup> Although either of the two-stage procedures provides an efficient approach for exploring the potential impact of covariates on group membership, they are not a substitute for a final analysis based on the jointly estimated model. Roeder, Lynch, and Nagin (in press) have found that the first described two-stage procedure in particular tends to overstate the statistical significance of covariates on group membership. The probable reason is that the procedure ignores the uncertainty about group assignments.

<sup>13</sup> Recent advances in latent curve modeling have adapted the conventional assumption of a continuous distribution of growth curves to accommodate the group-based approach described here. For an excellent summary of this advance in latent curve modeling, see Muthen (in press).

trajectories. Technically, the model is simply a mixture of probability distributions that are suitably specified to describe the data to be analyzed. Three such distributions were illustrated: the Poisson, which is suitable for analyzing count data; the censored normal, which is suitable for analyzing psychometric scale data with clusters of observations at the scale maximum or minimum or both; and the binary logit, which is suitable for analyzing binary data. For maximum flexibility the parameters that characterize the trajectory of each group are specified to vary freely across groups. This allows for substantial cross-group differences in the shape of trajectories.

The examples used to illustrate the method were selected to demonstrate four of the method's most important capabilities: (a) the capability to identify rather than assume distinctive developmental trajectories; (b) the capability to estimate the proportion of the population best approximated by the various trajectories so identified; (c) the capability to relate the probability of membership in the various trajectory groups to characteristics of the individual and his or her circumstances; and (d) the capability to use the posterior probabilities of group membership for various other purposes such as sampling, creating profiles of group membership, or serving as regressors in multivariate statistical analyses.

The modeling strategy described here has only recently been developed. Opportunities for extension abound. Three extensions seem particularly worthwhile. Further development of approaches for deciding on the best fitting number of groups would be extremely valuable. Although BIC scores are very useful for this purpose, in addition it would be helpful to have measures of the goodness of fit between actual and predicted trajectories. A second useful extension involves the joint estimation of trajectories of different behaviors—for example, joint estimation of psychological well-being in childhood and employment status over adulthood. This would provide the capacity to examine the relationship between the development patterns of distinct but likely linked behaviors. A third valuable extension involves developing diagnostics for calibrating the degree to which developmental trajectories cluster into distinct groups or vary regularly according to some specified parametric distribution (e.g., multivariate normal). Nagin and Tremblay (in press) described another rationale that is not discussed here for the group-based modeling strategy—to avoid making strong and generally untestable assumptions about the population distribution of de-

velopmental trajectories. The semiparametric mixture can be thought of as approximating an unspecified but likely continuous distribution of population heterogeneity in developmental trajectories. In so doing, a standard procedure in nonparametric and semiparametric statistics of approximating a continuous distribution by a discrete mixture is adopted (Follman & Lambert, 1989; Heckman & Singer, 1984; Lindsay, 1995). A difficult but valuable research effort would be to develop metrics for calibrating the degree to which trajectory groups are actually closely approximating a specified continuous distribution.

## References

- Appel, T. A. (1987). *The Cuvier–Geoffroy debate: French biology in the decades before Darwin*. New York: Oxford University Press.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioral research: Application and data analysis methods*. Newbury Park, CA: Sage.
- Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*, 29–53.
- Cantor, N., & Genero, N. (1986). Psychiatric diagnosis and natural categorization: A close analogy. In T. Millon & G. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 233–256). New York: Guilford Press.
- Caspi, A. (1998). Personality development across the life course. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 311–388). New York: Wiley.
- D'Unger, A., Land, K., McCall, P., & Nagin, D. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses of the London, Philadelphia, and Racine cohorts studies. *American Journal of Sociology*, *103*, 1593–1630.
- Erdfelder, E. (1990). Deterministic developmental hypotheses, probabilistic rules of manifestation, and the analysis of finite mixture distributions. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Time series and categorical longitudinal data* (Vol. 2, pp. 471–509). Boston: Academic Press.
- Farrington, D. P., & West, D. J. (1990). The Cambridge study in delinquent development: A prospective longitu-



- dinal study of 411 males. In H. Kerner & G. Kaiser (Eds.), *Criminality: Personality, behavior, and life history* (pp. 115–138). New York: Springer-Verlag.
- Follman, D. A., & Lambert, D. (1989). Generalized logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, *84*, 295–300.
- Ghosh, J. K., & Sen, P. K. (1985). On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. In L. M. LeCam & R. A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (Vol. 2, pp. 789–806). Monterey, CA: Wadsworth.
- Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.
- Greene, W. H. (1990). *Econometric analysis*. New York: Macmillan.
- Greene, W. H. (1991). LIMDEP User's Manual (Version 6.0) [computer software]. Bellport, NY: Econometrics Software, Inc.
- Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, *52*, 271–320.
- Holyoak, K., & Spellman, B. (1993). Thinking. *Annual Review of Psychology*, *44*, 265–315.
- Jones, B. L., Nagin, D. S., & Roeder, K. (1998). A SAS procedure based on mixture models for estimating developmental trajectories (Working Paper No. 684). Carnegie Mellon University, Department of Statistics.
- Kandel, D. B. (1975). Stages in adolescent involvement in drug use. *Science*, *190*, 912–914.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association*, *90*, 773–795.
- Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928–934.
- Keribin, C. (1997). *Consistent estimation of the order of mixture models* (Working Paper No. 61). Université d'Evry-Val d'Essonne, Laboratoire Analyse et Probabilité.
- Lambert, D. (1993). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*, 1–13.
- Land, K., McCall, P., & Nagin, D. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models with empirical applications to criminal careers data. *Sociological Methods & Research*, *24*, 387–440.
- Land, K., & Nagin, D. (1996). Micro-models of criminal careers: A synthesis of the criminal careers and life course approaches via semiparametric mixed Poisson models with empirical applications. *Journal of Quantitative Criminology*, *12*, 163–191.
- Lindsay, B. G. (1995). *Mixture models: Theory, geometry, and applications*. Hayward, CA: Institute of Mathematical Statistics.
- Loeber, R. (1991). Questions and advances in the study of developmental pathways. In D. Cicchetti & S. Toth (Eds.), *Models and integrations. Rochester Symposium on developmental psychopathology* (Vol. 3, pp. 97–115). Rochester, NY: University of Rochester Press.
- MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*, 215–253.
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110–133.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.
- Moffitt, T. E. (1993). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*, 674–701.
- Muthen, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*(4), 557–585.
- Muthen, B. O. (in press). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent curve modeling. In A. Sayers & L. Collins (Eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Association.
- Muthen, B. O., & Curran, P. (1997). General longitudinal modeling of individual differences in experimental design: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*, 371–402.
- Nagin, D., Farrington, D., & Moffitt, T. (1995). Life-course trajectories of different types of offenders. *Criminology*, *33*, 111–139.
- Nagin, D., & Land, K. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, *31*, 327–362.

- Nagin, D., & Tremblay, R. E. (in press). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*.
- Parzen, E. (1960). *Modern probability theory and its applications*. New York: Wiley.
- Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, *44*, 329–335.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–164.
- Raudenbush, S. W. (in press). Toward a coherent framework for comparing trajectories of individual change. In A. Sayers & L. Collins (Eds.), *New methods for the analysis of change*. Washington DC: American Psychological Association.
- Rindskopf, D. (1990). Testing developmental models using latent class analysis. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Time series and categorical longitudinal data* (Vol. 2, pp. 443–469). Boston: Academic Press.
- Roeder, K., Lynch, K., & Nagin, D. (in press). Modeling uncertainty in latent class membership: A case study from criminology. *Journal of the American Statistical Association*.
- Schwarz, G. (1978). Estimating dimensions of a model. *Annals of Statistics*, *6*, 461–464.
- STATA 4.0 [Computer software]. (1995). College Station, TX: STATA Corporation.
- Titterton, D. M., Smith, A. F. M., & Madov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24–36.
- Tremblay, R. E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The preschool behavior questionnaire: Stability of its factor structure between culture, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development*, *10*, 467–484.
- Wasserman, L. (1997). *Bayesian model selection and model averaging* (Working Paper No. 666). Carnegie Mellon University, Department of Statistics.
- Widiger, T. A., & Frances, A. (1985). The DSM-III personality disorders. *Archives of General Psychiatry*, *42*, 615–623.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381.

## Appendix

### Derivation of Likelihood

As described in the main text, the form of the likelihood for each individual  $i$  is

$$P(Y_i) = \sum_j \pi_j P^j(Y_i),$$

where  $P(Y_i)$  is the unconditional probability of observing individual  $i$ 's longitudinal sequence of behavioral measurements— $Y_i$ ,  $P^j(Y_i)$  is the probability of  $Y_i$  given membership in  $j$ , and  $\pi_j$  is the probability of  $j$ . Thus, the likelihood for the entire sample of  $N$  individuals is

$$L = \prod_{i=1}^N P(Y_i).$$

For given  $j$ , conditional independence is assumed for the sequential realizations of the elements of  $Y_i$ ,  $y_{it}$  over the  $T$  ages of measurement. Thus,

$$P^j(Y_i) = \prod_{t=1}^T p^j(y_{it}),$$

where  $p^j(y_{it})$  is the probability distribution function of  $y_{it}$  given membership in group  $j$ .

For the censored normal,  $p^j(y_{it})$  equals

$$p^j(y_{it} = S_{\min}) = \Phi\left(\frac{S_{\min} - \beta^j x_{it}}{\sigma}\right),$$

$$p^j(y_{it}) = \frac{1}{\sigma} \phi\left(\frac{y_{it} - \beta^j x_{it}}{\sigma}\right)$$

for  $S_{\min} \leq y_{it} \leq S_{\max}$ , and

$$p^j(y_{it} = S_{\max}) = 1 - \Phi\left(\frac{S_{\max} - \beta^j x_{it}}{\sigma}\right),$$

where  $\phi$  and  $\Phi$  are, respectively, the density function and cumulative distribution function of a normal ran-

dom variable with mean  $\beta^j x_{it} = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2$  and standard deviation  $\sigma$ , and  $S_{\min}$  and  $S_{\max}$  are, respectively, the scale minimum and maximum.

For the Poisson-based model,

$$p^j(y_{it}) = \frac{(\beta^j x_{it})^{y_{it}} e^{-\beta^j x_{it}}}{\beta^j x_{it}!}$$

Also, see Land, McCall, and Nagin (1996), Nagin and Land (1993), or Roeder, Lynch, and Nagin (in press) for a derivation of the likelihood for the still

more general form of the Poisson-based model that makes use of the zero-inflated Poisson distribution.

Finally, for the binary logit-based model,

$$p^j(y_{it}) = \frac{e^{\beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2}}{1 + e^{\beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2}}$$

Received June 15, 1998

Revision received October 16, 1998

Accepted January 21, 1999 ■

### Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time. If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.