Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method

Daniel S. Nagin
Carnegie Mellon University

Richard E. Tremblay
University of Montreal

This article presents a group-based method to jointly estimate developmental trajectories of 2 distinct but theoretically related measurement series. The method will aid the analysis of comorbidity and heterotypic continuity. Three key outputs of the model are (a) for both measurement series, the form of the trajectory of distinctive subpopulations; (b) the probability of membership in each such trajectory group; and (c) the joint probability of membership in trajectory groups across behaviors. This final output offers 2 novel features. First, the joint probabilities can characterize the linkage in the developmental course of distinct but related behaviors. Second, the joint probabilities can measure differences within the population in the magnitude of this linkage. Two examples are presented to illustrate the application of the method.

Two prominent themes in developmental psychology, developmental psychopathology, and developmental criminology are comorbidity and heterotypic continuity. Comorbidity refers to the contemporaneous occurrence of two or more undesirable conditions, such as conduct disorder and hyperactivity during childhood (Angold, Costello, & Erkanli, 1999; Nagin & Tremblay, 1999) or anxiety and depression in

adulthood (Kessler et al., 1994). Heterotypic continuity is the manifestation over time of a latent individual trait in different but analogous behaviors (Caspi, 1998; Kagan, 1969). For example, a propensity for violence may reveal itself as kicking and biting siblings during early childhood, gang fighting during adolescence, and spouse abuse during adulthood. The form and target of the aggression is different, but the constant is physical violence. Due to the changing form of the manifestation, use of the same measurement scale at different stages of life is inappropriate for capturing such a tendency.

Comorbidity and heterotypic continuity are typically represented by a single summary statistic, usually a correlation or odds ratio, that measures the co-occurrence of the two behaviors or symptoms of interest (e.g., hyperactivity and conduct disorder at age 6) or, alternatively, relates the two distinct behaviors measured at different life stages (e.g., physical aggression at age 5 and violent delinquency at age 15). Examples of research using this conventional measurement strategy in comorbidity analysis include work by Costello et al. (1988); Fergusson, Horwood, and Lynskey (1993); Haapasalo, Tremblay, Boulerice, and Vitaro (in press); Lewinsohn, Hops, Roberts, Seeley, and Andrews (1993); and Valez, Johnson, and Cohen (1989) and in heterotypic continuity analysis include Backteman and Magnusson (1981); Caspi (1998); Farrington (1990); Huesmann, Eron, Lefkowitz, and Walder (1984); Loeber and LeBlanc (1990); and Olweus (1979).

cd: 115

Figure 1 depicts the essence of the data summary problem in such analyses. Figure 1A characterizes the conventional approach to summarizing the co-occurrence of two behaviors. For a comorbidity analysis, up to T summary measures of association can be computed—one for each of the T measurement periods. For example, if series X and Y, respectively, measured depression and anxiety over T periods, up to T correlation coefficients could be computed to represent comorbidity for each period. For heterotypic continuity analysis, the combinations are potentially even larger because each of the T measurements of behavior X can be related to any of the K measurements of behavior Y from period T onward.

The conventional approach to measuring behavioral overlap and stability suffers from several important limitations. Most important, it makes inefficient use of longitudinal data because measures of association only use two assessment periods. This is especially problematic in light of the enormous cost of conducting longitudinal studies. Also, it is paradoxical because a key rationale for tracking individuals for more than two assessment periods is to provide the capacity to trace more than the linear change in the developmental course, yet this capacity is greatly dissipated by conventional two-period-based summary statistics for measuring comorbidity and heterotypic continuity. Second, the customary interpretation of a summary statistic is that its magnitude applies equally to all individuals within the population under study. For example, suppose the correlation between involvement in antisocial behavior at ages 8 and 14 was found to be 0.6. Most commonly, this correlation is interpreted as applying to all population members. However, there is a more complicated—and probably more realistic—alternative: It is an average correlation calculated over heterogeneous subpopulations. For some subpopulations, there may be very little association, whereas for
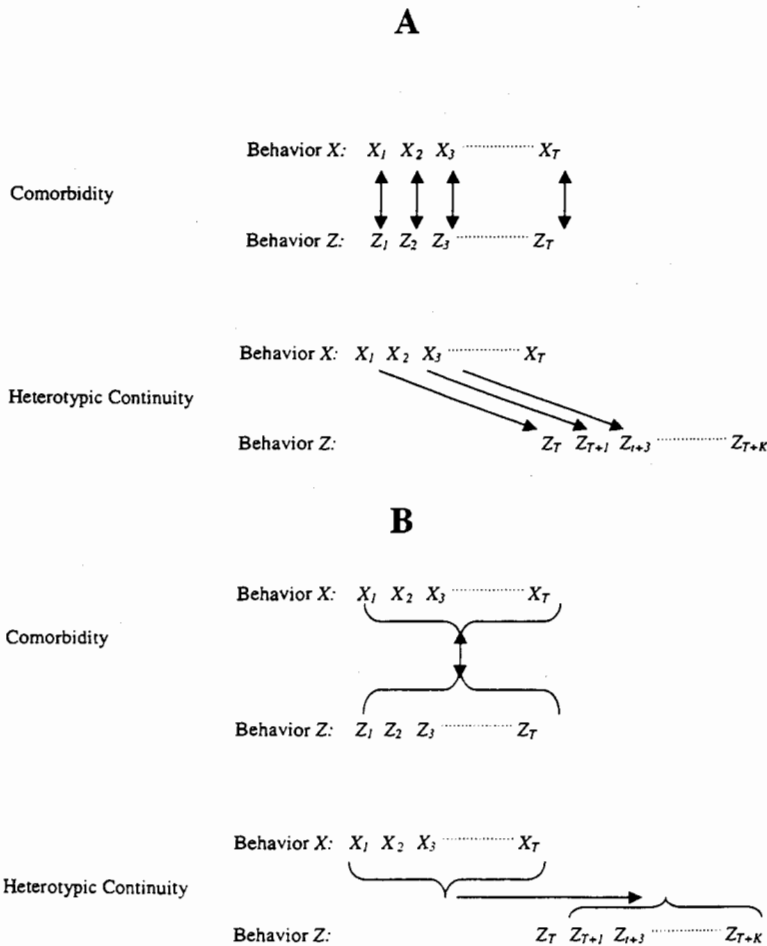

Figure 1. Measuring comorbidity and heterotypic continuity. A: Conventional approach. B: Joint trajectory approach.

other subpopulations, the association may be much larger (e.g., Magnusson & Bergman, 1990; Pulkkinen & Tremblay, 1992).

Two other limitations of the conventional approach are directly related to the use of correlation coefficients. The first is that the correlation's magnitude is not readily interpretable. For example, what does a correlation of +0.6 imply about the proportion of individuals who exhibit behavioral continuity compared to the proportion of individuals who exhibit behavioral change? The second limitation is that the correlation's size can be quite sensitive to the inclusion of outlier data from highly skewed distributions, such as those often encountered in research on mental health problems, crime, and antisocial behavior (Moffitt, 1993).

In this article, we develop and demonstrate a statistical model that relates the entire longitudinal course of the two behaviors of interest. The essence of the approach is depicted in Figure 1B. It aims to relate all measurements of the two behaviors of interest in a single summary statistical model. The model is a generalization of a group-based method described in Nagin (1999) for identifying distinctive groupings of individual-level trajectories within the population. Such trajectory groups describe the course of a behavior over age or time and might include "increasers," "decreasers," and "no-changers." The generalization provides the capacity for the joint estimation of trajectory models for two distinct but theoretically related measurement series.

The three key outputs of the joint model are illustrated in Figure 2 and Table 1. The table and figure contain examples of output from an analysis, to be developed in detail here, which examines the comorbidity of hyperactivity and physical aggression from ages 6 to 15 years in a large sample of boys from Montreal, Canada. One output is the identification of the optimal number of trajectory groups for each measurement series. Figure 2 depicts the form of the trajectory groups identified for these two externalizing behaviors. For each behavior, a four-group model was found to be optimal. A second key output is the probability of membership in each trajectory group. These probabilities are displayed in the first part of Table 1. For example, 3.9% of the population is estimated to belong to the chronic physical aggression trajectory group and 10.2% to a counterpart hyperactivity group. The third key output is the joint probability of membership in trajectory groups across behaviors. These probabilities are displayed in the second, third, and
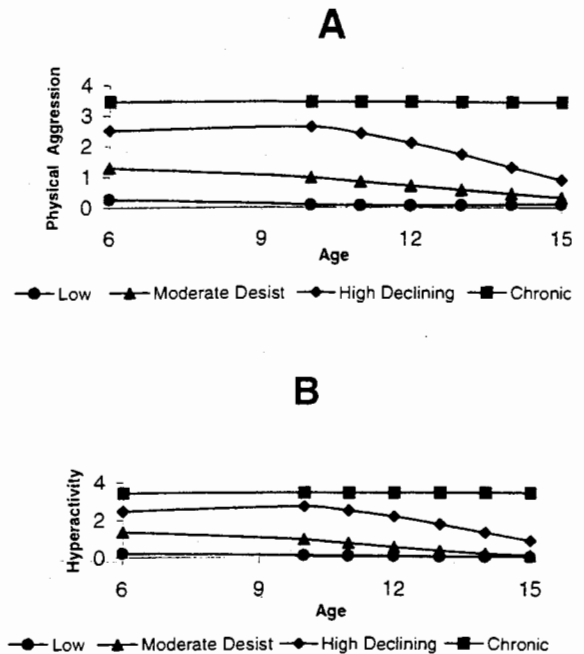

Figure 2. A: Trajectories of physical aggression: Ages 6 to 15. B: Trajectories of hyperactivity: Ages 6 to 15.

fourth parts of Table 1. For example, the estimated probability of belonging to the chronic physical aggression trajectory group given membership in the counterpart chronic hyperactivity group is .275 (second part of Table 1), whereas the converse conditional probability estimated is much higher, .722 (third part of Table 1). Finally, the joint probability of belonging to both chronic trajectory groups is .011 (fourth part of Table 1). These probabilities, each of which is important in its own right, provide the capacity to characterize the linkage in the developmental course of distinct but related behaviors.

The joint trajectory model advances conventional approaches to measuring comorbidity or heterotypic continuity by providing the capability to examine the linkage between the dynamic unfolding of the two behaviors over the entire period of observation. Thus, it makes use of all the longitudinal measurements of the behaviors of interest. In addition, it captures population differences in the strength and form of the comorbidity or heterotypic continuity.

Model

This section is intended to provide a conceptual description of the joint trajectory model. A detailed derivation of the likelihood function that is maxi-

Table 1
Comorbidity of Hyperactivity and Physical Aggression

Behavior or group	Low	Moderate desist	High declining	Chronic
Probability estimates for joint physical aggression and hyperactivity model				
Physical aggression	.283	.425	.253	.039
Hyperactivity	.220	.401	.277	.102
Probability of physical aggression group conditional on hyperactivity group				
Low hyperactivity	1.000	.000	.000	.000
Moderate desist hyperactivity	.158	.803	.038	.000
High declining hyperactivity	.000	.367	.594	.039
Chronic hyperactivity	.000	.013	.711	.275
Probability of hyperactivity group conditional on physical aggression group				
Low physical aggression	.777	.223	.000	.000
Moderate desist physical aggression	.000	.758	.239	.003
High declining physical aggression	.000	.062	.652	.286
Chronic physical aggression	.000	.000	.278	.722
Joint probability of hyperactivity and physical aggression trajectory group				
Low physical aggression	.223	.063	.000	.000
Moderate desist physical aggression	.000	.332	.101	.001
High declining physical aggression	.000	.016	.165	.072
Chronic physical aggression	.000	.002	.025	.011

mized for the purpose of model estimation is provided in the Appendix. First, the statistical model underlying the estimation of a group-based trajectory model for a single behavior is summarized. For an in-depth discussion of the estimation of the univariate trajectory model, see Nagin (1999). The approach used to link two univariate models to form a joint model is then described.

A developmental trajectory describes the developmental course of a behavior over age or time. The group-based model assumes that, as an approximation, the population is composed of a mixture of groups with distinctive developmental trajectories (Land, McCall, & Nagin, 1996; Nagin & Land, 1993; Nagin & Tremblay, 1999). However, at the level of the individual trajectory group, membership is not observed. Thus, individual level data cannot be sorted ex ante for the purpose of estimating each group's trajectory. Instead, the estimation procedure, which is based on mixture modeling, identifies the shape of the trajectory for each group and the proportion of the population that constitutes each such group.

The technical specifics of the statistical model used to identify and estimate the trajectory groups depend on the form of the response variable under investigation. For a count variable, it is assumed that the data are generated by an underlying Poisson process, in which the Poisson rate parameter for each group j comprising the population follows up to a third-order polynomial in age:

$$\log(\lambda_{it}^j) = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3,$$

where λ_{it}^j is individual i's rate of the behavior of interest at time t conditional on membership in group j, and Age_{it}, Age_{it}^2, and Age_{it}^3 are, respectively, subject i's age, age squared, and age cubed at time t.[1,2] The

[1] The model does not require that all individuals be the same age at each assessment period t. For example, the model can accommodate data from an overlapping cohort design. Thus, age is not synonymous with time. An alternative formulation of a trajectory is in terms of time. In this formulation, it is important that time be measured from a natural starting point like commencement of treatment for depression. Otherwise, a trajectory in time is not interpretable.

[2] As described in Jones, Nagin, and Roeder (in press), the Poisson-based model for count data is embedded in the more-general Zero-Inflated Poisson. As implied by its name, this generalization of the Poisson model includes a "zero-inflation" factor to account for the possibility of a greater frequency of zero realizations than is predicted by the standard Poisson distribution. Because our focus is on group-based modeling not on the analysis of count data, for

model's coefficients, β_0^j, β_1^j, β_2^j, and β_3^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.[3] This absence of constraint is key. It allows each group's trajectory to have a distinct shape. Thereby, the model has the capacity to capture unusual mixtures of developmental trajectories in the population.

For psychometric scale data, which is commonly censored at scale minimums and maximums and for binary data (e.g., accident or not in period t), we assume that the data for each group j are being generated by an underlying latent variable, y_{it}^{*j}, where

$$y_{it}^{*j} = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \epsilon_{it}$$

For the censored normal model, ϵ is a disturbance assumed to be independently normally distributed with zero mean and constant variance σ^2. For the binary model, it is assumed to follow the extreme value distribution, which underlies the logistic distribution (Maddala, 1983).

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

$$y_{it} = S_{min} \qquad \text{if } y_{it}^{*j} < S_{min},$$

$$y_{it} = y_{it}^{*j} \qquad \text{if } S_{min} \leq y_{it}^{*j} \leq S_{max},$$

and

$$y_{it} = S_{max} \qquad \text{if } y_{it}^{*j} > S_{max}.$$

For binary data, the observed quantity, which is again denoted by y_{it}, the model assumes that $y_{it} = 1$ if $y_{it}^{*j} > 0$ and equals 0 otherwise.

Key outputs of the univariate model are estimates of the coefficients defining the shape of the trajectory for each group j and also of π^j, the proportion of the population belonging to each group j. These are also included among the outputs of the joint model. The key additional output of the joint model is the estimation of probabilities linking the two behaviors under investigation.

Formally, let Y^1 and Y^2 denote the two longitudinal series of measurements for each individual i that are to be modeled in a joint trajectory format, where Y^1 is measured over T^1 periods, Y^2 is measured over T^2

periods, and the index i designating the individual has been suppressed for notational convenience. Note that the model does not require that measurements be contemporaneous or that the length of the measurement period be the same (i.e., $T^1 \neq T^2$). For example, in an illustration of heterotypic continuity described below, oppositional behavior and property delinquency were measured over two age intervals of differing length and were also partially overlapping: 6–15 years old and 11–17 years old, respectively. Moreover, the form of the data need not be the same for the two series. The joint model can accommodate any combination of censored normal, count, or binary data across the two measurement series under investigation.

Figure 1B suggests two conceptual models for joining the trajectories for Y^1 and Y^2. For heterotypic continuity, we are interested in linking an earlier behavior, denoted by x (e.g., childhood opposition) with a later behavior denoted by z (e.g., adolescent property delinquency). Let j and k index the trajectory groups associated with behaviors x and z, respectively. Formally, we are interested in estimating the conditional probability of transiting to trajectory k (e.g., a trajectory of chronic property offending in adolescence) given membership in trajectory group j (e.g., a trajectory of chronic opposition in childhood). This probability is denoted by $\pi^{k|j}$.

For comorbidity analysis, the arrows linking behaviors x and z are double headed to suggest an interest in measuring their joint occurrence. Angold et al. (1999) emphasize the importance of measuring both $\pi^{k|j}$ (e.g., the probability of chronic childhood physical aggression given chronic childhood hyperactivity) and the converse conditional probability $\pi^{j|k}$ in describing comorbidity. In addition, the joint probability of belonging to both trajectory groups (e.g., following trajectories of chronic hyperactivity and chronic physical aggression) is of interest. This probability is denoted by π^{jk}.

As it turns out, these two alternative representations of the joint occurrence of trajectory groups are analytically equivalent. Specifically, the likelihood function can be specified such that the direct output of estimation are either estimates of (a) π^{jk} for $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$ or (b) π^j and $\pi^{k|j}$ for $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$.[4] These two alter-

ease of exposition we use the more familiar Poisson distribution.

---

[3] In principle, any order polynomial of age can be used to model λ_{it}^{*j}, subject to identifiability.

[4] As discussed in the Appendix, an equivalent version of the second parameterization is in terms of π^k and $\pi^{j|k}$.

natives are formally equivalent because if we know π^{jk}, we can calculate π^j, $\pi^{k|j}$, π^k, and $\pi^{j|k}$ as follows:

$$\pi^j = \sum_k \pi^{jk}, \qquad j = 1, \ldots, J, \qquad (1)$$

$$\pi^{k|j} = \frac{\pi^{jk}}{\pi^j}, \qquad (2)$$

$$\pi^k = \sum_j \pi^{jk}, \qquad k = 1, \ldots, K, \qquad (3)$$

and

$$\pi^{j|k} = \frac{\pi^{jk}}{\pi^k}. \qquad (4)$$

Alternatively, if we know π^j and $\pi^{k|j}$, we can calculate π^{jk}, π^k, and $\pi^{j|k}$ as follows:

$$\pi^{jk} = \pi^{k|j}\pi^j, \qquad (5)$$

$$\pi^k = \sum_j \pi^{k|j}\pi^j, \qquad k = 1, \ldots, K, \qquad (6)$$

and

$$\pi^{j|k} = \pi^{k|j}\frac{\pi^j}{\pi_k}. \qquad (7)$$

We bring these calculations to the reader's attention not only to make clear that the "comorbidity" and "heterotypic" models are analytically equivalent models but also because the calculations themselves can be usefully employed in analysis; a point we hope to demonstrate in the next section where we provide illustrative applications. In both these illustrations, we use the second, "heterotypic continuity" parameterization in estimation and thus use Equations 5–7 to calculate other probabilities of interest.

Two Illustrative Examples

In this section, the application of the joint models is illustrated with two examples. The first is a joint trajectory model of the comorbidity of hyperactivity and physical aggression. The second is a joint trajectory linking childhood opposition to adolescent property delinquency. An SAS-based procedure for estimating joint models such as these is available by request. This procedure extends the univariate trajectory estimation software described in Jones, Nagin, and Roeder (in press).

Both examples use data from a Montreal-based pro-spective longitudinal study. This study tracks 1,037 white males of French ancestry. Subjects were selected in 1984 from kindergarten classes in low-socioeconomic Montreal neighborhoods. Following the assessment at age 6, the boys and other informants were interviewed annually from ages 10 to 17. Assessments were made on a wide range of factors. Among these were physical aggression, hyperactivity, and opposition, which were measured at age 6 and annually from ages 10 to 15, based on teacher ratings using the Social Behavior Questionnaire (Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987). Physical aggression was assessed with three items: kicks, bites, hits other children; fights with other children; and bullies or intimidates other children. Its minimum and maximum scores are, respectively, 0 and 6. Hyperactivity was assessed with two items: squirmy, fidgety; does not keep still. Its minimum and maximum scores are 0 and 4. Opposition, which ranges from a score of 0 to 10, was measured by teacher ratings of five items: does not share materials; irritable; disobedient; blames others; and inconsiderate. The property offense scale, which was based on self-reports of activity over the past year, included the following items: stealing from a store, keeping objects worth more than $10, stealing something worth more than $100, entering without paying admission, stealing money from home, stealing a bicycle, stealing something worth between $10 and $100, buying stolen goods, being in an unauthorized place, and breaking and entering (Nagin & Tremblay, 1999).

Example 1: Comorbidity Analysis

Much research has documented the overlap of physical aggression and hyperactivity in children (Angold et al., 1999; Hinshaw, Lahey, & Hart, 1993; Hinshaw, Zupan, Simmel, Nigg, & Melnick, 1997; Kerr, Tremblay, Pagani-Kurtz, & Vitaro, 1997; Lahey, McBurnett, & Loeber, in press; Tremblay, Mâsse, Pagani, & Vitaro, 1996). This example is intended to illustrate how a joint trajectory analysis can illuminate the nature of this overlap by measuring it from a dynamic perspective.

Figure 2 reports a joint trajectory model of the comorbidity of physical aggression and hyperactivity among the subjects of the Montreal study. Figure 2A and 2B report the trajectory model for physical aggression and hyperactivity, respectively. For both of these psychometric scales, the censored normal distribution is used to model the trajectories to account for the censoring at the lower and upper bounds of the

scale. A discussion of model selection, the optimal number of groups and the order of the polynomial characterizing each group, follows in a separate section.

The trajectories for hyperactivity and physical aggression are very similar. For both behaviors, there is a group called "lows," whose individuals rarely display the behavior. A second group is best characterized as moderate-level desisters. At age 6, they manifested modest levels of the behavior, but by ages 10–12 they have largely desisted from displays of that behavior. A third group, called "high-level decliners," start off scoring relatively high on the problem behavior at age 6, but by age 15 score far lower. Finally, there is a group of chronics that start off scoring high on the behavior and continue to score high throughout the observation period.

The first part of Table 1 reports estimates of the probability of membership in each trajectory group. As previously noted, one such set of marginal probabilities estimates, $\hat{\pi}^j$, emerges directly from estimation. Estimates of the marginal probabilities for the other behavior, $\hat{\pi}^k$, can be calculated by Equation 6 based on the estimates of $\pi^{k|j}$ and π^j produced directly from model estimation. For both hyperactivity and physical aggression, group membership probabilities are about the same: 20%–30% for the lows, approximately 40% for the moderate-level desisters, 25%–30% for high-level decliners, and 5%–10% for the chronics.

Thus far, the analysis suggests that hyperactivity and physical aggression follow nearly identical developmental courses. However, the comorbidity probabilities reported in Table 1 suggest the overlap in their developmental courses is not nearly complete. Specifically, two sets of conditional probabilities are reported: (a) the probability of membership in each of the physical aggression trajectories conditional on membership in a given hyperactivity trajectory group (second part of Table 1), and (b) the converse set of probabilities for each hyperactivity group conditional on a given physical aggression group (third part of Table 1). In addition, estimates of the joint probabilities of trajectory group membership are reported in the fourth part of Table 1. The probabilities in the third part of Table 1 were calculated with Equation 7 and those in the fourth part with Equation 5.

Both sets of conditional probabilities show a high level of overlap in similar counterpart trajectory groups. This is suggested by the generally large diagonal elements of the probability matrices in the second and third parts of Table 1, which indicate substantial comorbidity. Still there are important differences, particularly for the extreme groups. Nearly 100% of those in the low-hyperactivity group are estimated to belong to the low-physical-aggression group. By contrast, it is estimated that 22% of those in the low-physical-aggression group belong to the moderate hyperactivity group. Thus, while virtually all low-hyperactivity boys are also low-physical-aggression boys, included among the low-physical-aggression boys is a sizeable minority of moderately hyperactive boys. This suggests that an observation of low hyperactivity is synonymous with low physical aggression, but that the reverse is not necessarily true. Some children who are not physically aggressive may nonetheless display modest hyperactivity. Similarly, when considering the chronic trajectories, only 28% of the chronic-hyperactive group are also members of the chronic-physical-aggression group. By contrast, 72% of the chronic-physical-aggression group are estimated to belong to the chronic-hyperactive group. Thus, most chronically physically aggressive boys are also chronically hyperactive; however, the reverse is not necessarily true—most chronically hyperactive boys are not chronically physically aggressive. Combined, these results suggest that the overlap between physical aggression and hyperactivity is complex. Low hyperactivity predicts low physical aggression better than low physical aggression predicts low hyperactivity. However, at high levels of these behaviors the reverse is true. Chronic physical aggression better predicts chronic hyperactivity than chronic hyperactivity predicts chronic physical aggression.

The fourth part of Table 1 reports the joint probabilities of trajectory group membership. Due to the high comorbidity of these two externalizing behaviors, most of the probability mass is concentrated on the diagonal elements of the matrix. The modal group is composed of those belonging to the two moderate desisting trajectory groups. This group accounts for 33.2% of the population. In total, 61.8% belong to trajectory groups that are low or moderate on both behaviors. In contrast, only 1.1% are estimated to follow a joint trajectory of both chronic physical aggression and chronic hyperactivity.

Example 2: Heterotypic Continuity Analysis

In this example, a joint trajectory model is estimated to examine the heterotypic continuity between opposition and property offending. As discussed in

Nagin and Tremblay (1999), a key question in developmental criminology is whether there is a single pathway to all criminal behavior (e.g., Gottfredson & Hirschi, 1990) or multiple pathways. Loeber and colleagues (Loeber, 1991; Loeber et al., 1993) propose a multiple pathway model in which different childhood developmental pathways lead to different types of offending in adolescence and adulthood. One such developmental pattern is a covert behavior-problem pathway that starts with minor problems such as lying, which leads to property damage, and is then followed by serious covert delinquent acts, such as fraud and burglary. The joint model of opposition and property delinquency is intended to explore this linkage.

Specifically, this analysis relates trajectories of opposition from ages 6 to 15 to trajectories of property delinquency from ages 11 to 17. Figure 3A displays four trajectories of opposition. Accompanying probabilities of group membership are reported in Table 2. The opposition trajectory groups are very similar to those for hyperactivity and physical aggression both in terms of their shapes and group membership probabilities: a low group, a moderate desisting group, a high declining group, and a chronic group that

are respectively estimated to account for 25%, 46%, 24%, and 5% of the population. Figure 3B reports six distinct trajectories of property delinquency. One group, which we call lows, comprise about 30% of the population. This group has a rate of offending that is near zero over the entire observation period, which indicates that they engage in virtually no property delinquency. Three of the group trajectories are rising steadily. One of these, the rising chronics, is small in size, an estimated 5.9% of the population, but follows a very interesting trajectory for delinquency researchers—it starts high and rises steeply to a very high level. The other two, the low-rising 1s and 2s, start near zero, with the low-rising 2s rising to a distinctly higher level than do the low-rising 1s. Both of these are examples of late-onset property delinquency. Combined, they are estimated to comprise 41.3% of the population. The remaining two trajectory groups display greater stability, although both show some clear evidence of decline. One group, the "medium decliners," accounts for an estimated 14.6% of the population. The other group, labeled "medium chronics," comprise an estimated 7.3% of the population.

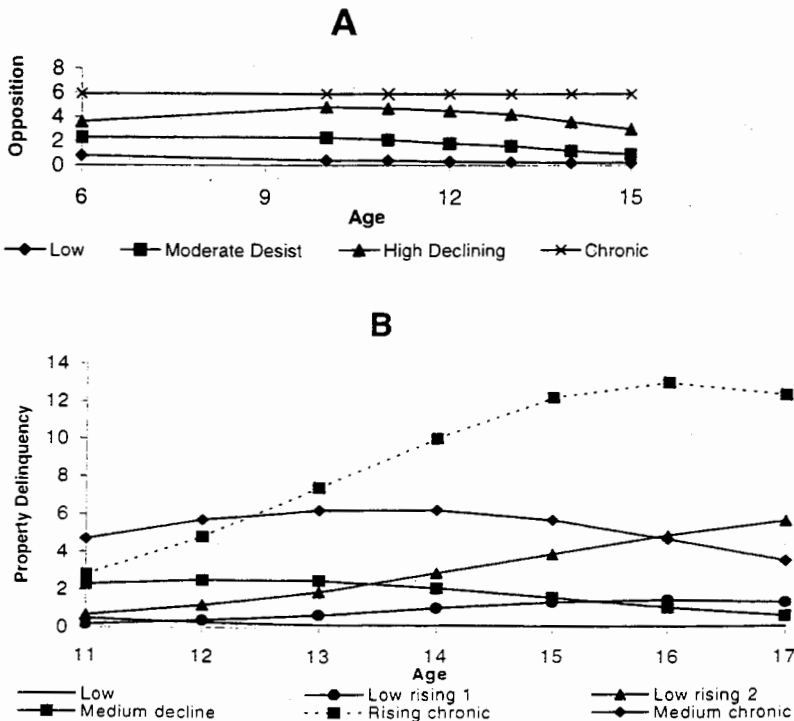Table 2 reports transition probabilities linking these

Figure 3. A: Trajectories of opposition: Ages 6 to 15. B: Trajectories of property delinquency: Ages 11 to 17.

Table 2
Property Delinquency Group Membership Probabilities Conditional on Opposition Group Membership

Opposition group	Property delinquency group					
	Low	Low-Rising 1	Low-Rising 2	Medium declining	Rising chronic	Medium chronic
Low	.532	.259	.084	.093	.009	.023
Moderate desist	.274	.280	.173	.158	.051	.064
High declining	.164	.222	.191	.189	.110	.124
Chronic	.204	.245	.164	.087	.160	.139

two sets of developmental trajectories.[5] Observe that a majority of the low-opposition boys are in the low-property delinquency trajectory group (.532), whereas the probability of a boy in the chronic group following this trajectory is only .204. By contrast, the chronic-opposition group has a combined probability of .299 (= .139 + .160) of following the rising chronic or medium-chronic trajectories, whereas for the low-opposition group this probability is only .03. Thus, as argued by Loeber and colleagues (1993), opposition has a clear linkage to property offending. Still, the link falls far short of certainty. Nearly half of the low-opposition boys follow one of the five more-elevated property delinquency trajectories, and more than 10% follow the three most elevated trajectories—the low-rising 2, rising chronic, or medium-chronic trajectories. On the other hand, not all the chronic boys engage in high levels of property delinquency—about 45% follow the low or low-rising 1 trajectories.

Model Selection

One of the most technically challenging problems in mixture modeling is determining the optimal number of groups to include in the model. One possible choice for testing the optimality of a specified number of groups is the likelihood ratio test. The likelihood ratio test is only suitable for model selection problems in which the alternative models are nested linear subspaces. In mixture models, a k group model is not a nested linear subspace of a $k + 1$ group model, and, therefore, it is not appropriate to use the likelihood ratio test for model selection (Erdfelder, 1990; Ghosh & Sen, 1985; Titterington, Smith, & Makov, 1985). Instead, use of the Bayesian Information Criterion (BIC) is recommended as a basis for selecting the optimal model (D'Unger, Land, McCall, & Nagin, 1998). For a given model, BIC is calculated as

$$BIC = \log(L) - 0.5 * \log(n) * (k),$$

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) argue 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 recommend selection of the model with the maximum BIC. In this application, model selection includes not only the determination of the number of groups, but also the order of the polynomial used to capture the shape of each trajectory group.

There is, however, one practical complication in applying BIC to model selection in the joint trajectory format. The number of models that must be estimated grows with the square of the number of models that are considered under the univariate trajectory format. If N^1 and N^2 models are considered for Y^1 and Y^2, respectively, an exhaustive model search requires estimating $N^1 * N^2$ joint models. Instead, it is recommended that model selection be based on searches of the two univariate model spaces, which thereby reduces the number of models searched to $N^1 + N^2$. The final joint model is estimated with the number and shapes of trajectories found to be optimal, based on the two univariate model searches. Experience has shown that trajectories emerging from joint estimation differ little from their univariate counterparts.

Application of the maximum BIC criterion for model selection generally leads to a clear-cut choice of the best model. For example, BIC-based calculations reported in Nagin (1999) strongly support the four-group physical aggression model shown in Fig-

---

[5] It is also possible to calculate the probability of membership in each opposition trajectory group conditional upon membership in each delinquency group. These probabilities are not reported because they seem less useful given the temporal sequencing of the behaviors.

ure 2 being the "best" model. However, in some applications, the BIC score continues to improve as more groups are added. In this circumstance, the addition of a new group to the model most commonly results in the splitting of a large group into two smaller groups with parallel trajectories. This in fact occurred for the property delinquency trajectory analysis reported in Figure 3. We stopped at six groups because the addition of more groups only split the large groups of low delinquency boys, but left the high delinquency trajectory groups unchanged. Given that the latter were of greatest interest, the addition of more groups was not informative.

Confidence Intervals for Group Membership Probabilities

An alternative to the joint model for joining trajectories of different measurement series relies only on the univariate trajectory method. This can be accomplished as follows: (a) estimate separate univariate trajectory models for Y_1 and Y_2, (b) based on the posterior probabilities of group membership, sort the estimation sample members into the trajectory groups that they most likely belong to for Y_1 and Y_2; and (c) cross-tabulate the group membership counts to estimate $\pi^{k|j}$, $\pi^{j|k}$.[6] This approach has two shortcomings that are avoided by the joint estimation procedure. First, as described in Roeder, Lynch, and Nagin (1999), such a "classify-analyze" strategy will not produce consistent estimates of the above probabilities. Second, it provides no valid basis for computing the standard errors of the estimates of $\pi^{k|j}$, $\pi^{j|k}$, and π^{jk}. Consequently, it is not possible to calculate confidence intervals for the probability estimates or to conduct hypothesis tests about the estimates. In contrast, the joint model provides consistent and efficient estimates of all required standard errors.

The only obstacle to calculating confidence intervals with the joint model is a practical one. Specifically, one must ensure that the estimates of π^j and $\pi^{k|j}$ remain in the 0–1 interval. This is accomplished by estimating π^j and $\pi^{k|j}$ as multinomial logistic probabilities:

$$\pi_j = \frac{e^{\theta_j}}{\sum_j e^{\theta_j}}$$

and

$$\pi_{k|j} = \frac{e^{\gamma_{k|j}}}{\sum_k e^{\gamma_{k|j}}}, \qquad j = 1, \ldots, J,$$

where θ^j and $\gamma^{k|j}$ are parameters to be estimated.

The downside of this approach is that the estimates of π^j and $\pi^{k|j}$ are nonlinear functions of the parameter estimates. Therefore, confidence intervals for π^j and $\pi^{k|j}$ are not directly computable by the usual calculations based on the standard errors of θ^j and $\gamma^{k|j}$. Two alternatives are available for calculating confidence intervals. One can use the first term of a Taylor Expansion on the logistic equation above to form an equation that is linear in the estimated parameters (Greene, 1990); or one can use the parametric bootstrap technique, which was used here, to estimate the standard errors of $\hat{\pi}^j$ and $\hat{\pi}^{k|j}$. This procedure, which was first proposed by Efron (1979), simulates the sampling distribution of $\hat{\pi}^j$ and $\hat{\pi}^{k|j}$ as follows: (a) Drawing on the result that $\hat{\pi}^j$ and $\hat{\pi}^{k|j}$ are asymptotically multivariate normally distributed with mean and covariance (π^j, Σ^j) and $(\pi^{k|j}, \Sigma^{k|j})$, respectively, and that the points estimates themselves are consistent estimates of the population quantities, a simulated random sample of 10,000 estimates of θ^j and $\gamma^{k|j}$ is generated; (b) these estimates are then used to generate 10,000 estimates of π^j and $\pi^{k|j}$, which, in turn, are rank ordered to create simulated sampling distributions of $\hat{\pi}^j$ and $\hat{\pi}^{k|j}$. Thus, a 95% confidence interval is simply the lower and upper 2.5 percentiles of this distribution. More generally, this same approach can be used to calculate confidence intervals for other more-complicated nonlinear functions such as $\hat{\pi}^j \hat{\pi}^{k|j} = \hat{\pi}^{jk}$.

Table 3 illustrates the application of this approach to the opposition-property delinquency trajectory model. Point estimates and 95% confidence intervals are reported for three probabilities of potential interest to researchers. The first is the probability of belonging to the chronic-opposition trajectory group, π^3_{op}. The point estimate of this quantity is .047 with a 95% confidence interval of .021–.101. Note, unlike the typical confidence interval, this bootstraped interval is asymmetric. This asymmetry makes sense because the point estimate of π^3_{op}, .047, is close to the lower theoretical bound of a probability, 0; therefore, there

---

[6] Following model estimation, the parameter estimates can be used to calculate the probability that each individual's observed pattern of behavior was generated by each trajectory group. These probabilities are called the posterior probabilities of group membership.

Table 3
Confidence Intervals for Various Probabilities: Opposition–Property Delinquency Model

Probability	Calculation	Point estimate	95% confidence interval
Chronic opposition	$\hat{\pi}_{op}^3 = \dfrac{e^{\vartheta^4}}{1 + e^{\vartheta^2} + e^{\vartheta^3} + e^{\vartheta^4}}$	.047	.021–.101
Chronic or high desister opposition	$\hat{\pi}_{op}^3 + \hat{\pi}_{op}^4$	.290	.239–.353
Chronic opposition and rising chronic property or moderate chronic property	$\hat{\pi}_{Del}^{5|4}\hat{\pi}_{op}^4 + \hat{\pi}_{Del}^{6|4}\hat{\pi}_{op}^4$	.014	.005–.033

is more upside error potential. Also reported is the probability and associated confidence interval for the linear sum of two probabilities, the probability of belonging to the chronic or high-desister group. This probability is estimated at .290 with 95% confidence interval of .239–.353. The third probability quantity is a joint probability—the probability of belonging to the chronic childhood-opposition group and then proceeding to belong to one of the two most theft-prone groups in adolescence, the rising- and medium-chronic groups. Like the first example this probability is small, .014, and consequently, its 95% confidence interval, .005–.033, is substantially asymmetric. Each of these illustrative confidence intervals suggests that despite the large number of parameters estimated in a joint model, it is still possible to obtain reasonably precise estimates of important probability estimates.

Strengths and Limitations of a Group-Based Modeling Approach

Our statistical depiction of comorbidity and heterotypic continuity requires a group-based modeling approach. Without it, we would not be able to probabilistically link the developmental course of two different behaviors. Use of probabilities to characterize alternative developmental courses requires the designation of discrete states of the behaviors—developmental trajectories. In this respect, our modeling strategy has its roots in the tradition of characterizing comorbidity and heterotypic continuity with odds ratios rather than correlations. Use of the odds ratio also requires the creation of behavioral states such as highly anxious or a diagnosis of depression.

As discussed in Nagin (1999), Nagin and Land (1993), and Nagin and Tremblay (1999), the assumption that the population is composed of distinct groups is unlikely to be strictly correct. Instead, the groups are intended as an approximation of an underlying

continuous process. In so doing, we adopt a standard procedure in nonparametric and semiparametric statistics of approximating a continuous distribution from a discrete mixture (Follman & Lambert, 1989; Heckman & Singer, 1984; Lindsay, 1995; Manski, 1995).

This statistical approximation has its analog in the taxonomic-based theories that are quite common in developmental psychology. 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). 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 distinct groups.

The idea of using groups to approximate a continuous distribution is easily illustrated with an example. Suppose Figure 4A depicts the population distribution of some behavior z. In Figure 4B, this same distribution is replicated and overlaid with a histogram that approximates its shape. This panel illustrates that any continuous distribution with finite endpoints can be approximated by a discrete distribution composed of a finite number of "points of support" (i.e., the shaded "pillars"). For any given number of points of support, maximum likelihood estimation can be used to estimate two sets of parameters. The first identifies the location on the horizontal axis of each point of support. In Figure 4, these points are denoted by $\bar{z}_1, \bar{z}_2, \bar{z}_3,$ ...., where $\bar{z}_j$ measures the "average" behavior of in-
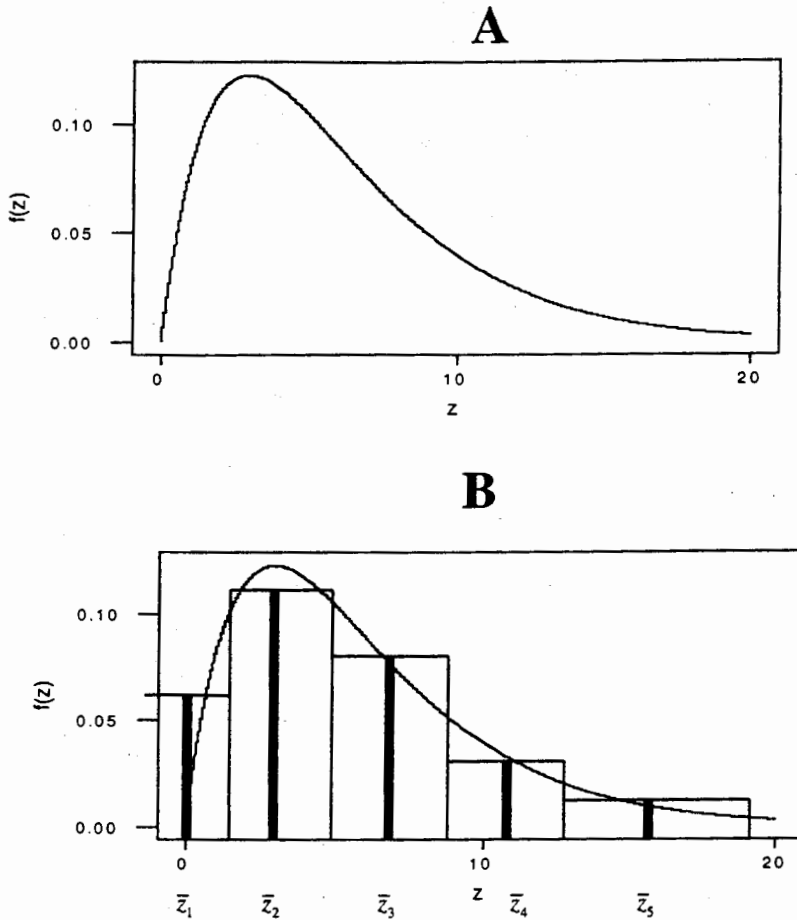
A

B

Figure 4. Using groups to approximate a continuous distribution. A: An unknown continuous distribution. B: An unknown continuous distribution approximated by a finite number of points of support.

dividuals at the *j*th point of support. A second set of parameters measure the proportion of the population π_j at each point of support. These proportions must sum to 1 but in general will not be equal. If a time-age dimension were added to Figure 4 to measure the developmental trend of $\bar{z}_j$, each of these points of support would correspond to the trajectory groups depicted. For example, in Figures 2 and 3, the estimates of π_j would correspond to the proportion of the population whose developmental trajectory is best approximated by group *j*.

When is a group-based approach a useful approximation of an underlying continuous phenomenon? The answer to this question depends on many issues but we emphasize two here. First, if the behavior under study is reasonably approximated by a known continuous distribution that is also tractable (e.g., the multivariate normal distribution), an analysis based on that distribution is obviously the prefer-

red approach. Alternatively, if a suitable continuous distribution is not known or tractable, the group-based, semiparametric approach is an attractive alternative.

A second and related rationale for group-based modeling is that it is well suited for the study of a behavior that does not vary regularly throughout the population, but instead tends to reveal itself in markedly different intensities in clusters of individuals. Figure 5 displays two contrasting possibilities. In Figure 5A, the behavior varies uniformly across the population, whereas in Figure 5B there are two distinct modes of behavior. A group-based modeling approach makes little sense for the behavior depicted in Figure 5A—there are no distinct groups. However, for the behavior in Figure 5B, it makes a great deal of sense because of the bimodality of the behavior. A two-group model would capture the clusters at each mode.
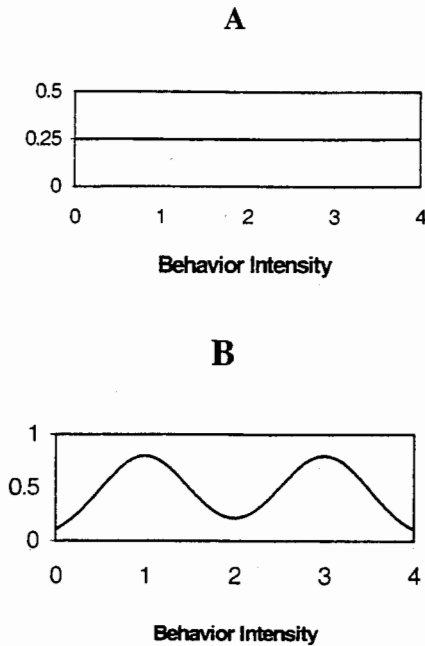
A

Behavior Intensity

B

Behavior Intensity

Figure 5. Two hypothetical examples of the population distribution of a behavior. A: Uniformly distributed behavior. B: Bimodally distributed behavior.

The examples in Figure 5 are in one dimension. In one dimension, clear-cut evidence of clustering requires unusual examples such as bimodality. However, trajectories of development are in two dimensions—Behavior × Time. In this circumstance, the possibilities for distinctive groups of trajectories are not hard to imagine. As previously noted, developmental psychology is filled with examples of theories predicting distinctive patterns of development. We make this point because the real world is mostly composed of phenomena in which clustering or regular variation is not clear-cut, particularly if a time dimension is a component of the analysis.

Our experience with model selection for the property delinquency illustrates this ambiguity. On the one hand, the group-based strategy was successful in identifying distinctive trajectories of serious property delinquency that would be difficult to identify with modeling approaches based on continuous distributions. These groups are the two-dimensional analog of the Figure 5B distribution. On the other hand, the group-based modeling strategy successively split moderate-level delinquency trajectory groups into similar parallel groups. For these groups, trajectories varied more regularly in the population. Such groups are the two-dimensional analog to the Figure 5A distribution. This example illustrates that the choice be-

tween the multinomial modeling strategy that underlies the group-based approach and the modeling strategy based on a multivariate continuous distribution will rarely be clear cut ex ante. Only the analysis can reveal whether variation is highly regular or clustered into distinctive groups or both.

The successive splitting of the moderate-level delinquency groups also provides a concrete illustration of a fundamental statistical problem that attends to using a finite number of support points to approximate a continuous distribution. Perfect conformance to the distribution requires an infinite number of points of support. Thus, even though statisticians have made important progress in demonstrating the utility of the BIC in identifying the optimal number of groups in mixture problems, this theoretical work begins with the assumption that the population is composed of a finite, albeit unknown, number of groups.

Thus, it must be recognized that unless one holds to the position that the population is strictly composed of discrete groups, use of any goodness-of-fit measure, BIC, or otherwise, will not formally identify the correct model. We do not hold the view that the population is strictly composed of distinct groups. One practical consequence of this view is that as sample size increases, models will tend to include more groups because ever more information will become available for identifying more subtle features of the response surface.[7] However, in our view, this is not a fatal shortcoming of semiparametric or nonparametric modeling strategies such as that used here. Rather, it is a frank acknowledgment that *all* models are approximations and thus, literally speaking, incorrect. Still more work is needed for developing methods to calibrate the adequacy of group-based models. Such methods should include not only mathematical statistical criteria but also visual representations of the response surface.

One of psychometrics' most important contributions to statistical theory was the development of methodology for linking test items responses to latent constructs such as intelligence. Structural equation modeling has its roots in this tradition. The modeling strategy used here and the predecessor work on uni-

---

[7] Note, however, that D'Unger et al. (1998) found that the optimal number of groups as determined by BIC was largely insensitive to sample size. This finding reflects the fact that BIC's penalty for adding more parameters grows with sample size.

variate trajectory estimation are not in this tradition. Instead, it is rooted in the classical statistical and econometric tradition in which the quantity to be measured is the quantity of interest. Thus, in the model section, the term *latent variable* was used to describe y_{it}^{*j} because it is not fully observed. This use of the term "latent" is different from that in the psychometric literature, in which the term *latent factor* refers to an unobservable construct that is assumed to give rise to multiple manifest variables. Thus, within the formal structure of the model, we do not attempt to formally link the response variable to some more fundamental but not directly measured construct.

The primary advantage of this approach is that it helps streamline the statistical model. This is a critical advantage because the model is already complex. Furthermore, we suspect that adapting this modeling framework to a structural equation modeling framework would make it difficult to retain two key strengths of the framework—the flexibility to handle a variety of different data types and to accommodate missing data.

Balanced against these advantages are obvious limitations. The modeling framework provides no formal basis for combining items intended to be indicators of an unobserved latent construct and does not provide the capacity to explicitly account for measurement error. For an excellent account of the application of structural equation modeling methods to group-based modeling, see Muthén and Muthén (1999).

Conclusion

This article has demonstrated a group-based method for joining developmental trajectories of distinct but theoretically related behaviors. The objective was to provide a method that improves on conventional approaches for measuring comorbidity and heterotypic continuity. The principal advantages of this approach are as follows: (a) It links the entire developmental course of the two behaviors of interest rather than relating single measurements of each behavior made at a particular time—in so doing, the method makes use of all the data assembled over the course of the longitudinal study; (b) it provides three readily interpretable metrics for describing comorbidity and heterotypic continuity, π^{klj}, π^{jlk}, and π^{jk}; and (c) the group-based method is specifically intended for problems in which theory or practice suggests distinctive developmental courses within the population.

As the examples illustrate, the joint model is well suited for assessing population heterogeneity in the form of the comorbidity or heterotypic continuity under study.

A particularly valuable next step in the development of the joint model methodology is generalizing the specification of π^{klj} to incorporate covariates. This generalization has already been demonstrated for the univariate model (Nagin, 1999; Roeder et al., 1999). The obstacle to adding covariates to the joint model is developing a workable approach to limit the number of parameters to be estimated. Consider the opposition-property delinquency model, which includes four opposition trajectories and six delinquency trajectories. Without any constraints on the specification of π^{klj}, 20 ($= 4 * (6 - 1)$) parameters would be required for each covariate. This is plainly too many, and actual experience confirms this prediction—the standard errors of the estimated parameters are extremely large. Thus, as a practical matter, the fully general model is not identified, except perhaps in extremely large data sets. A much more constrained but still sensible model is required.

References

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40,* 57–87.

Backteman, G., & Magnusson, D. (1981). Longitudinal stability of personality characteristics. *Journal of Personality, 49,* 148–160.

Brame, B., Nagin, D., & Tremblay, R. (in press). Developmental trajectories of physical aggression from school entry to late adolescence. *Journal of Clinical Psychology and Psychiatry.*

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* (pp. 311–388). New York: Wiley.

Costello. E. J., Costello, A. J., Edelbrock, C., Burns, B. J., Dulcan, M. K., Brent, D., & Janiszewski, S. (1988). Psychiatric disorders in pediatric primary care: Prevalence and risk factors. *Archives of General Psychiatry, 45,* 1107–1116.

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.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1–26.

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. (1990). Childhood aggression and adult violence: Early precursors and later-life outcomes. In D. J. Pepler and K. H. Rubin (Eds.), *The development of childhood aggression.* Hillsdale, NJ: Erlbaum.

Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1993). Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15 year olds. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 1127–1134.

Follman, D. A. & Lambert, D. (1989). Generalized logistic regression by non-parametric 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 and R. A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer: Vol. II.* (pp. 789–806). Monterey, CA: Wadsworth.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford, CA: Stanford University Press.

Greene, W. H. (1990). *Econometric analysis.* New York: Macmillan.

Haapasalo, J., Tremblay, R. E., Boulerice, B., & Vitaro, F. (in press). Predicting preadolescents' delinquency, social withdrawal and school placement from kindergarten behavior: Person and variable approaches. *Quantitative Criminology.*

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica, 52,* 271–320.

Hinshaw, S. P., Lahey, B. B., & Hart, E. L. (1993). Issues in taxonomy and comorbidity in the development of conduct disorder. *Development and Psychopathology, 5,* 31–49.

Hinshaw, S. P., Zupan, B. A., Simmel, C., Nigg, J. T., & Melnick, S. (1997). Peer status in boys with and without attention-deficit hyperactivity disorder: Prediction from overt and covert antisocial behavior, social isolation, and authoritative parenting beliefs. *Child Development, 68,* 800–896.

Holyoak, K., & Spellman, B. (1993). Thinking. *Annual Review of Psychology, 4,* 265–315.

Huesmann, L. R., Eron, L. D., Lefkowitz, M. M., & Walder,

L. O. (1984). Stability of aggression over time and generations. *Developmental Psychology, 20*(6), 1120–1134.

Jones, B., Nagin, D. S., & Roeder, K. (in press). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Research & Methods.*

Kagan, J. (1969). The three faces of continuity in human development. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 983–1002). Chicago, IL: Rand McNally.

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.

Kerr, M., Tremblay, R. E., Pagani-Kurtz, L., & Vitaro, F. (1997). Boys' behavioral inhibition and the risk of later delinquency. *Archives of General Psychiatry, 54,* 809–816.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Study. *Archives of General Psychiatry, 51,* 8–19.

Lahey, B. B., McBurnett, K., & Loeber, R. (in press). Are Attention-Deficit/Hyperactivity Disorder and Oppositional Defiant Disorder developmental precursors to Conduct Disorder? In M. Lewis & A. Sameroff (Eds.), *Handbook of developmental psychopathology.* New York: Plenum Press.

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.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology, 102,* 133–144.

Lindsay, B. G. (1995). *Mixture models: Theory, geometry, and applications.* Hayward, CA: The 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.

Loeber, R., & LeBlanc, M. (1990). Toward a developmental criminology. *Crime and Justice, 12,* 375–437.

Loeber, R., Wung, P., Keenan, K., Giroux, B., Southamer-

Loeber, M., Van Kammen, W. B., & Maughan, B. (1993). Developmental pathways in disruptive child behavior. *Development and Psychopathology, 5,* 103–133.

Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics.* New York: Cambridge University Press.

Magnusson, D., & Bergman, L. R. (1990). A pattern approach to the study of pathways from childhood to adulthood. In L. N. Robins & M. Rutter (Eds.), *Straight and devious pathways from childhood to adulthood* (pp. 101–116). New York: Cambridge University Press.

Manski, C. F. (1995). *Identification problems in the social sciences.* Cambridge, MA: Harvard University Press.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Moffitt, T. E. (1993). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674–701.

Muthén, B., & Muthén, L. (1999). *Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes.* Manuscript in preparation.

Nagin, D. (1999). Analyzing developmental trajectories: Semi-parametric, group-based approach. *Psychological Methods, 4,* 139–177.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology, 31,* 327–362.

Nagin, D. S., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70,* 1181–1196.

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin, 86,* 852–875.

Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist, 44,* 329–335.

Pulkkinen, L., & Tremblay, R. E. (1992). Patterns of boys' social adjustment in two cultures and at different ages: A longitudinal perspective. *International Journal of Behavioural Development, 15*(4), 527–553.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–164.

Roeder, K., Lynch, K., & Nagin, D. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association, 94,* 766–776.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* New York: Wiley.

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.

Tremblay, R. E., Mâsse, L. C., Pagani, L., & Vitaro, F. (1996). From childhood physical aggression to adolescent maladjustment: The Montreal Prevention Experiment. In R. D. Peters & R. J. McMahon (Eds.), *Preventing childhood disorders, substance abuse and delinquency* (pp. 268–298). Thousand Oaks, CA: Sage.

Valez, C. N., Johnson, J., & Cohen, P. (1989). A longitudinal analysis of selected risk factors of childhood psychopathology. *Journal of American Academy of Child and Adolescent Psychiatry, 28,* 861–864.

(Appendix follows)

Appendix

Specification of Likelihood Function

This appendix specifies the form of the likelihood function used to estimate the joint trajectory model. Given that the joint model is an elaboration of the univariate model, we begin with a brief summary of the form of its likelihood function. For each individual i, the likelihood of observing the longitudinal sequence of behavioral measurements, $Y_i = \{y_{i1}, y_{i2}, y_{i3}, \ldots, y_{iT}\}$, denoted by $P(Y_i)$, is specified as a mixture of J underlying latent groups:

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

where $P^j(Y_i)$ is the probability of Y_i given membership in group j and π^j is the probability of group j. As described in the main text, the form of $P^j(Y_i)$ is selected to conform with the type of data under analysis. In work to date, the Zero-Inflated Poisson distribution has been used in the analysis of count data, the censored normal distribution in the analysis of psychometric data, and the logit distribution in the analysis of binary data. See Jones et al. (in press) or Nagin (1999) for a full discussion of these alternatives. The basic model also assumes that conditional on membership in group j, the random variables y_{it} are independent. Thus,

$$P^j(Y_i) = \prod^T P^{jt}(y_{it}).$$

The assumption of conditional independence is a strong one, but not nearly as strong as it may seem at first blush. As described in the main text, the parameters of the polynomial defining each trajectory group are allowed to vary freely. Thus, in the population, individual-level behavior will be correlated over time even though within-group, it is independent over time. Thus, the model does conform with the empirical reality that at the population level there is serial correlation in behavior. Further, we note that the standard fixed or random effect model generally assumes that conditional on the fixed or random effect (and other relevant covariates), individual-level behavior is independent over time. In principle, the conditional independence assumption could be relaxed to allow for within-group serial correlation in behavior, but the cost in terms of complexity would be considerable.

The joint trajectory model builds from the univariate model as follows: Let Y^1 and Y^2 denote the two longitudinal series to be modeled in a joint trajectory format, where Y^1 is measured over T^1 periods, Y^2 is measured over T^2 periods, and the index i has been suppressed for notational convenience. We continue the maintained assumption of conditional independence given group membership. Thus,

$$f^j(Y^1) = \prod^{T_1} f^{jt}(y_t^1)$$

and

$$h^j(Y^2) = \prod^{T_2} h^{jt}(y_t^2),$$

where $f(*)$ and $h(*)$ are suitably defined probability distributions given the form of the data.

We next add a new layer to the conditional independence assumption with the assumption that conditional on j and k, Y^1 and Y^2 are independently distributed, $P^{jk}(Y^1, Y^2) = f^j(Y^1)h^j(Y^2)$. Thus, the unconditional likelihood function of Y^1 and Y^2 sums across $P^{jk}(Y^1, Y^2)$, with each such conditional distribution weighted by the joint probability of membership in trajectory group j for Y^1 and trajectory group k for Y^2, π^{jk}:

$$P(Y^1, Y^2) = \sum_j \sum_k \pi^{jk} f^j(Y^1) h^j(Y^2).^{A1}$$

An alternative and equivalent form of the likelihood function builds from the result that $\pi^{jk} = \pi^{k|j}\pi^j$. Thus,

$$P(Y^1, Y^2) = \sum_j \sum_k \pi^{k|j}\pi^j f^j(Y^1)h(^k Y^2)$$

$$= \sum_j \pi^j f^j(Y^1) \sum_k \pi^{k|j} h^k(Y^2).$$

The reader will observe that this second likelihood function has a sequential construction—each group j of Y^1 is linked to each group k of Y^2 via a conditional probability $\pi^{k|j}$. For problems in which Y^1 temporally precedes Y^2, this formulation is natural. However, regardless of temporal sequence, still another equivalent formulation conditionally links each group k to each group j via the conditional probability $\pi^{j|k}$. For this formulation, the likelihood function for each individual i is:

$$P(Y^1, Y^2) = \sum_k \pi^k \left[h^k(Y^2) \sum_j \pi^{j|k} f^j(Y^1) \right].$$

[A1] Brame et al. (in press) use a distinctly different approach. Instead of linking trajectories across Y^1 and Y^2 via π^{jk}, the univariate approach is expanded by estimating trajectories groups that combine parameters for each behavior. Specifically in this formulation a group is defined by three sets of parameters, π^m, $\mathbf{B}^{1m}$, and $\mathbf{B}^{2m}$, where m indexes the combined trajectory, π^m is the proportion of the population in each combined group, and $\mathbf{B}^{1m}$ and $\mathbf{B}^{2m}$ are vectors of parameters specifying the shape of group m's trajectory for behaviors Y^1 and Y^2, respectively. For this formulation, the form of the likelihood function for a given individual i is $P(Y^1, Y^2) = \sum_m \pi^m f^m(Y^1)h^m(Y^2)$.