Criminal Careers: Discrete or Continuous?

David F. Greenberg¹

Received: 17 August 2015 / Revised: 1 March 2016 / Accepted: 10 March 2016 /
Published online: 28 March 2016
© Springer International Publishing AG 2016

Abstract

Purpose Numerous empirical studies of criminal careers have made use of finite mixture modeling to analyze sequences of events such as crimes or arrests. This paper aims to demonstrate that the analysis of criminal careers can benefit from the use of alternative methods, including multilevel methods, and individual time series.

Methods We use multilevel nonlinear modeling and individual time series techniques to analyze artificial data as well as arrest histories for 3432 males released from the California Youth Authority in 1981 and 1986, and followed for several decades after release.

Results Multilevel methods are capable of identifying discrete groups in longitudinal data. In the California Youth data set, we find little clear evidence of sharply discrete arrest trajectories.

Conclusions We recommend that researchers explore alternatives to finite mixture modeling when analyzing criminal career data.

Keywords Criminal careers · Finite mixture modeling · Individual time series · Latent growth curve modeling

Mixture Modeling of Criminal Career Trajectories

The classification of criminals into discrete criminal types was an important element of late nineteenth-century positivist criminology [8–10, 35, 96], but until recently, the appeal of general theories of crime causation (e.g., strain theory, learning theory, control theory) left the enterprise of constructing taxonomies somewhat marginal to mainstream twentieth-century criminology. It has, however, been revived in the past 40 years, with a refocused interest on criminal careers—that is, the sequence of offenses or arrests committed by law violators over time.

✉ David F. Greenberg
dg4@nyu.edu

¹ Sociology Department, New York University, 295 Lafayette St., 4 Floor, New York, NY 10012, USA

The ground for this revival was paved by Chicago School sociologists whose qualitative investigations focused on the processes by which youths were initiated into delinquent activities through social interactions (Shaw 1930, 1931; [7]). Around the same time, Glueck and Glueck [37, 38] tracked the criminal involvement of Boston youths over time, and carried out quantitative analyses of the patterns they found. Perhaps because of the difficulty in following large numbers of young people over an extended period of time, this type of criminological research was never widely adopted. The studies of self-reported delinquency undertaken in the 1960s and 1970s were usually cross-sectional.

Wolfgang et al. [124] revived the longitudinal study of criminality with their research on the arrest histories of Philadelphia boys. Soon thereafter, criminologists began to develop statistical models describing the temporal patterns disclosed in those arrest histories [14, 16]. In this new body of work, it is the classification of careers, rather than the classification of criminals on the basis of static individual traits, that is at the center of attention.

Statistical analyses of temporal patterns of involvement in criminal conduct have been the subject of a number of books, and numerous journal articles. Theory and research on this topic carried out in the 1970s dealt with such issues as the shape of the age-crime curve, which rises in childhood, peaks in adolescence or early adulthood, and then declines [31, 40, 41, 51, 52]. This pattern was, Hirschi and Gottfredson [51, 52] proclaimed, the same for blacks and whites, males and females, and in different times and places—a contention that others disputed [41, 43, 109].

In retrospect, it is striking that the debate over this claim largely focused on the comparison of aggregate figures—that is, on the total numbers of arrests or convictions committed by persons of different ages, while having little to say regarding individual variability about the aggregate trends. The lack of long-term longitudinal data sets of individual crime records forced researchers of the late twentieth century to work with data for aggregates. In the interim, information about individual involvement in crime spanning a number of years has become available, making it possible to study individual patterns of criminality temporally. This availability has enabled researchers to overcome another weakness in the earlier body of work: it drew inferences about the effect of age from cross-sectional data, potentially confounding age effects with cohort effects [44].

The heated debates about the universality of the age-crime curve have largely been resolved. The existence of diverse patterns of temporal change in individual offenders' involvement in crime¹ is well-established [15, 27, 29, 43, 78, 109]. Research has also identified life events that alter patterns of criminality [13, 59, 60]. Nevertheless, despite vigorous research efforts spanning several decades, some fundamental questions about criminal careers remain unanswered. This paper addresses one of them—the shape of the distribution across individuals of criminal career trajectories. In particular, is this

---

¹ Technically speaking, most of the studies analyze actions taken by the criminal justice system, such as arrests or convictions—not crimes. Distributions of arrests and convictions can be shaped by the responses of criminal justice agents and organizations to criminal events. This complicates the interpretation of results, as it cannot be assumed that trajectories for these official actions coincide with those for criminal conduct. For present purposes, I ignore this complication, and will refer to crime trajectories when discussing arrests or convictions. In substantive research, it could be important to take into account the possibility of differential law enforcement practices based on race, ethnicity, sex and age, for example.

distribution discrete or continuous? Much of the recent research on criminal careers has adopted a particular strategy for studying temporal patterns of crime—the estimation of finite mixture models, also known as group-based trajectory analysis or latent class analysis² [2, 74–77, 84]. In a review of this research published just 8 years ago, Piquero [94] tallied more than 80 criminological publications using this approach. A more recent review located 105 studies [54]. Sterba et al. [110] observe that hundreds of studies using this method have been carried out in psychology.

There are no doubt numerous reasons for the popularity of finite mixture modeling as a strategy for studying criminal careers; one of them is that it overcomes an important limitation to earlier statistical methods for analyzing longitudinal patterns. Older methods could analyze heterogeneity in criminal career trajectories due to sources of heterogeneity that were known to researchers and measured. They could do this using interaction terms or through subgroup analyses. Yet researchers commonly do not know all of the sources of heterogeneity. Finite mixture modeling allows researchers to take into account heterogeneity due to causes that are not known to the researcher, and that are not represented by variables in the data set.

Although some writers have employed other statistical methods for studying criminal careers, such as multilevel modeling [17, 29, 34, 53, 81, 115], these alternatives have been used much less often than finite mixture models. This is so even though several researchers have expressed doubts or misgivings about the finite mixture approach³ [4–6, 26, 28, 97, 101, 102, 107, 110], and even though these methods can also address heterogeneity due to unknown and unmeasured causes. Finite mixture modeling clearly dominates research on this topic. The present paper is intended to clarify some of the issues that have figured in the debates about the potential value of this particular statistical tool to the study of criminal careers, and more generally, to developmental studies. It is also intended to encourage researchers to consider the full range of methodological options available to them, and to offer guidance to researchers considering these options as to how they might best choose models for their analyses.

The main statistical modeling approaches currently being used for analyzing criminal careers strive to represent data regarding criminal events in simplified form by positing a simple functional dependence of those events on time or on an individual's age. They then seek to model individual variability in the parameters that define these functions. The methods do this in different ways. In both the multilevel modeling and the finite mixture modeling approaches, this dependence is usually taken to be a polynomial of second or third degree. In most instances, a polynomial of low degree will provide only an imperfect fit to the sequence of criminal events that make up a career. The trajectory is thus a mathematical construct that the researcher employs to approximate the sequence of actual events [72, 75]. Rather than assuming that a single functional dependence holds exactly for the entire population of offenders (the standard assumption in an OLS or Poisson regression), the multilevel approach and the finite mixture approach both allow for individual variability in the parameters that characterize this functional dependence. Multilevel modeling assumes a normal distribution for

² Finite mixture modeling is not restricted to longitudinal analyses; it can also be applied to cross-sectional regressions [61]. When one considers the volume of research on criminal careers conducted using the group-based approach, the paucity of applications to cross-sectional analyses in criminology is astonishing.

³ There have also been responses to the critics, e.g., Nagin and Tremblay [83, 84] and Brame et al. [19].

the unmeasured random effects characterizing individual differences. In contrast, the finite mixture modeler assumes that the complete set of individual sequences can be treated as realizations of a finite number of discrete latent trajectories, each with its own trajectory parameters. Here, the distribution assumed for the unmeasured sources of heterogeneity is taken to be discrete, but not otherwise specified. The procedure estimates the parameters characterizing each latent trajectory, along with the probabilities that a given individual is following each trajectory. One of the appeals of the group-based approach is that it need not assume that the individual effects are normally distributed. Most of the time, researchers have no strong reason for thinking the individual effects to be Gaussian, so the ability to avoid reliance on an uncertain assumption is attractive.

It is common practice in the group-based approach to assign individuals to the most likely trajectory, given the observed sequence of offenses in that individual's criminal history. This method treats all the trajectories in a given set as being the same⁴ [2, 68, 113]. Consequently, variation in the trajectories being studied is modeled as being entirely due to membership in the classes [58]. The strategy groups together subjects whose sequences of criminal events are fairly similar, and places in separate groups subjects whose sequences are dissimilar. Models of this sort can be estimated in Stata (Partha Deb's *fmm* routine and Bobby Jones's *traj*), SAS's PROC TRAJ [55], Latent Gold [118], MPlus [71, 72], and R [11, 62]. In Stata's *fmm* routine, the researcher proceeds by estimating a model on the assumption that there are two discrete groups, then three, then four, and so forth. Fit statistics, including the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are used, along with several other criteria, to determine the number of groups that provide an optimum fit to the observed sequences of offenses in the sample⁵ [74, 75, 87]. Researchers studying criminal careers in this way commonly conclude that optimal fits can be obtained with three to seven groups [27, 29, 30, 54, 64, 84, 95, 112].

Statisticians, including the developers of the finite mixture modeling approach, have noted a limitation of the method: it does not test the assumption that the distribution being modeled consists of a discrete number of groups. When a finite mixture model is estimated, the estimation procedure fits the data *on the assumption that the assumed number of groups is correct*. It will fit the data as closely as possible whether or not the distribution being fit actually consists of discrete groups [18, 19, 75, 76, 110]. That this is so raises a legitimate concern that if a distribution does not actually contain discrete groups, the algorithms used in finite mixture modeling could, nevertheless, specify an optimal number of discrete groups that is larger than 1.

Whether the over-extraction of groups is a problem depends on the purpose of the analysis. If the empirical distribution of events under study consists of a mixture of a finite number of discrete groups, and one of the purposes of the analysis is to determine

⁴ Kreuter and Muthén [58] have recently synthesized the finite mixture model and hierarchical linear modeling approach in a manner that allows this assumption to be relaxed.

⁵ The models with different numbers of groups are not nested. Consequently, the likelihood ratio chi-square test does not have its usual distribution. For this reason, the AIC and BIC statistics are used in model selection, with most researchers opting for the BIC criterion. The smaller the AIC or BIC statistic, the better the fit, given the sample size and the number of parameters being estimated. Simulations lead Brame et al. [18] to conclude that in some circumstances, AIC performs better, while in others, BIC is preferable. Additional research on the merits of different fit measures can be found in [63, 87].

how many groups are present, then one will want to determine this number accurately. If the distribution is actually continuous, with no sharply demarcated groups, and one wants to know this, then the failure of the method to determine this will be equally troubling. There are other purposes, however, where the finite mixture modeling may be perfectly satisfactory even if it does not perfectly capture all of the features of the underlying distribution of trajectories. This paper is primarily concerned with analyses in which the researcher wants to know the shape of the true underlying distribution—in particular, whether it is continuous or discrete. Toward the end, the paper will also offer some brief remarks about analyses where this is not the goal.

A modest number of studies have examined the ability of finite mixture modeling to capture the features of known artificial data sets, or to compare it with multilevel modeling [18, 20, 21, 63, 73, 107, 110, 117]. To date, however, only a limited range of data patterns have been examined. Though there are particular circumstances where each of the two methods outperforms the other, in many circumstances encountered in research, the multilevel approach outperforms the group-based approach, even in the presence of moderate departures from normality [110].

Further simulations carried out by Warren et al. [121] add to concerns about the group-based approach. Comparing six different implementations of the basic idea, these authors found that they did not agree on the optimum number of groups to include in a model, and could not be counted on to identify the correct number. They could not reliably predict the group to which a subject belonged, or the proportion of subjects associated with a given trajectory, or the qualitative features of the trajectories estimated. As the authors point out, these simulations were carried out on ideal data sets, with no missing values and no sample attrition. In less than ideal circumstances, they observe, performance might be worse. Other researchers have found the optimal number of groups and the assignment of individuals to groups to depend on the length of the follow-up [28].

Simulations conducted by Schork and Schork [103] and Bauer and Curran [4] are of particular interest in relation to our investigation. Their studies were designed to assess the possibility that finite mixture models will yield optimum fits for models with more than one group even when the data-generating process does not have distinct groups, merely because of skewed distributions in the data. Their simulations show that when there is just one true group, the group-based approach will generally yield one group when the parameters are normally distributed,⁶ but will “find” more than one group when the parameters are skewed.

I confirmed their conclusions by generating 1000 observations for random draws from the beta distribution $B(a, b, x)$, with the shape parameters $a=2$, $b=6$. The beta distribution is a family of continuous distributions whose shape parameters are defined by the formula for the probability density

$$B(a, b, x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \quad (1)$$

when x lies in the closed interval $[0,1]$, and is 0 otherwise. This theoretical distribution is shown in Fig. 1a, while a histogram of the distribution of the random draws from that

⁶ I confirmed that Stata's *fmm* routine does, in fact, perform in this way.

Fig. 1 a Graph of Beta(2, 6). b Histogram of 1000 random draws from a B(2,6) distribution

distribution is displayed in Fig. 1b. Clearly, there are no distinct groups here. By construction, we have a unimodal, smooth, moderately skewed distribution, apart from irregularities originating in the randomness of the draw.

To assess the performance of finite mixture modeling in relation to distributions with these features, I carried out *fmm* estimation in versions 11 and 14 of Stata, assuming that the data originate in a mixture of 1, 2, 3, or 4 groups, each of which has a normal probability density distribution. The optimum fit is the

Table 1 BIC values (finite mixture estimation of 1000 random draws from a B(2,6) distribution)

Number of groups	Mix of normals	Mix of student's t	Mix of lognormal	Mix of gamma
1	-1084.97			
2	-1162.77	-1118.47	-1191.87 ^a	-1208.20 ^a
3	-1177.13 ^a	-1148.28 ^a	-1184.79	-1194.33
4	-1163.38	-1135.64	-1184.79	
5	-1145.29	-1137.31		

^a Starred item is lowest in each column. The 4-group lognormal model would not converge

model with the smallest value of BIC.⁷ The BIC statistics displayed in Table 1 prefer the model with three groups as optimal.⁸ If, instead of assuming that we have a mixture of normal distributions, we assume that we have a mixture of Student's t distribution, the optimum number remains 3. If we assume a mixture of lognormal or gamma distributions, the optimum number becomes 2.

The proportions of cases belonging to each group of the three-group mixture-of-normals model are shown Table 2. The proportion of cases for which a given trajectory is most likely are all substantial enough that we would want to retain all three groups in our model. Table 3 shows the Bayes' factor appropriate for comparing the probability that the correct model is i with the probability that it is j .⁹ Using Jeffrey's interpretation of these factors, in a comparison of the three-group model with the two-group model, the estimated Bayes' factor of 1311.60 signifies extremely strong evidence for three groups compared to two; in a comparison of the three-group model with the two-group model, the Bayes' factor of 19.01 also provides very strong evidence for three instead of four. The probability that the three-group model is correct, using the criterion of Kass and Wasserman [57]; see also [56, 75, p. 70], is 0.80.

Several criteria have been proposed to assess the model one has chosen. Nagin [75] recommends the computation of the average assignment probabilities based on the rule of assigning each case to the group that has the largest posterior probability. Tables 4 displays these averages (AvePP). If all cases were being assigned by the maximum posterior probability rule to the right group, the averages would be 1 for the proper group. Here, they are .713 for those assigned to the first group, .856 for those assigned to the second group, and .756 to the third group. Nagin [75] expresses a "personal rule of thumb" that all groups should have AvePP values of .70 or higher. Assignments for this model all meet this standard.

⁷ Note that Stata defines the BIC statistic as -2 times the value given by the formula in Nagin [75: 64].

⁸ The *fmm* routine does not permit estimation for a model with just one group. However, one can do a one-group estimation using the regular Stata command for the model, for the complete sample. As a check, I repeated the Stata estimation using the regression routine in version 4 of Latent Gold. The optimum number of groups was again 3, with the value of the BIC statistic, and the probabilities for each group relatively insensitive to the choice of the Bayes constant. I was unable to carry out this computation for the one-group model assuming a gamma or lognormal distribution because the software does not estimate models with these distributions. The Stata routine *traj* also fails to offer these modeling options.

⁹ Using Stata's convention for defining the Bayes Information Criterion, the formula is $p/p = \exp[(BIC - BIC)/2]$. This formula is predicated on the assumption that priors are uninformative.

Table 2 Proportions of cases in each class (three group model, mixture of normals) (finite mixture estimation of 1000 random draws from a B(2,6) distribution)

	Group		
	1	2	3
Proportion in class based on estimated posterior probability	.098	.335	.567
Proportion in class based on most likely probability	.123	.375	.502

A second criterion Nagin [75] recommends for assessing a model's adequacy is the odds of correctly classifying a case into the proper group based on the maximum probability rule, compared to the odds of doing so by assigning cases randomly in the proportions estimated to exist in the population. For our three groups, the *OCC* statistics are, respectively, 11.51, 11.29, and 3.17. The higher the *OCC* value, the higher the model's assignment accuracy. Nagin considers values greater than 5 for all groups to indicate high classification accuracy. We easily surpass this standard for groups 1 and 2, and fall somewhat short for the third group. If assignments to groups were being made at random, the odds of an assignment being correct would be right 49.4 % of the time. With the model, the assignment would be right 75.6 % of the time. Not bad!

A third criterion available for assessing the usefulness of a model is how well it separates the groups. Do they overlap a little, or a lot? Table 5 addresses this question by showing the estimated means of the three groups, along with their standard errors. It is clear that the three-group model separates the three groups quite effectively. None of the 95 % confidence limits (LCL and UCL) overlap.

The entropy is another measure of how useful a model is in classifying cases into categories. It can take on values between 0 (useless) and 1 (maximally useful). The three-group model has an entropy of .533, which is respectable but not spectacular.

Most published criminological studies of trajectories using the group-based approach rely on the BIC criterion, along with the principle that models should not have trajectories followed by very few individuals. Employing these criteria, our artificial example clearly has more than one group, with the precise number considered optimal dependent on the shape of the distributions we posit for the components of the mixture. A researcher considering the additional criteria as well would probably conclude that these models are acceptable. Yet these groups are artifacts of the algorithm used in the computation. They do not demonstrate that more than one group is present. (We will say more about these results at a later point).

In a more elaborate simulation, I generated trajectory models in which the values of an interval-level outcome variable measured annually on a sample of individuals between ages 12 and 41 were characterized by a quadratic dependence on age with coefficients being random draws from a normal distribution. I again obtained optimal

Table 3 Bayes factors (B_{12}) for pairs of groups (three group model, mixture of normals) (finite mixture estimation of 1000 random draws from a B(2,6) distribution)

Groups	(B_{ij})
3 with 2	1311.60
3 with 4	19.01

Table 4 Average assigned probabilities based on highest posterior probability rule (finite mixture estimation of 1000 random draws from a B(2,6) distribution)

Assigned group	Number assigned	Group		
		1	2	3
1	178	.713	.030	.257
2	283	.000	.856	.144
3	539	.063	.180	.756

fits for the coefficients and slopes with models that had more than one group. These results show that estimation results from a finite mixture model can be untrustworthy guides to the true number of groups actually present in a data set.

The findings of the Bauer and Curran [4] study and of the present simulations are especially troubling in relation to criminal career research because, in many populations, the distribution of offenses is highly skewed. This is just the circumstance in which the group-based approach is known to over-estimate the number of groups present in a data set. Because of the possibility that the research findings based on finite mixture model methods applied to genuine data sets in previous criminal career research are artifacts of the methods used to analyze the data, their findings as to the number of groups present in a data set cannot be assumed, without further investigation, to reflect the underlying distribution of the parameters characterizing the individual subjects of the study. Indeed, critics of the finite mixture modeling approach have observed that if there are, in fact, no discrete groups, analyses carried out on the assumption that they exist could be misleading [97]. Our analyses demonstrate the perspicuity of this observation, and motivate the exploration of alternatives to the group-based procedures.

Theoretical Expectations About Groups

This paper takes it to be substantively interesting to know whether the distribution of criminal careers is discrete or continuous. The answer to this question may bear on our theoretical understanding of the processes that generate or inhibit criminal conduct. For this reason, it may be true but dissatisfying simply to observe that a continuous distribution can be approximated to an arbitrary degree of accuracy with a finite sum of discrete distributions [5]. That does not answer the question.

Table 5 Parameter estimates of three group model (3-group model, mixture of normals) (finite mixture estimation of 1000 random draws from a B(2,6) distribution)

Group	Coefficient	Std. Error	z	Prob.	LCL	UCL
Group 1	.080	.011	7.63	.000	.060	.101
Group 2	.378	.027	14.16	.000	.326	.431
Group 3	.207	.013	15.79	.000	.181	.233

Theoretical considerations might lead us to anticipate that criminal careers are discrete. We expect that people tend to form friendships on the basis of homophily [69]. It is a well-known feature of social life that those with similar values and lifestyles tend to like and seek each other out. They will tend to dislike those whose values and activity rounds differ sufficiently from their own, and will avoid socializing with them. Through verbalized support and rewards for behaviors that distinguish the group, as well as through imitation, these patterns of differential association will reinforce the group differences established by initial patterns of friendship selection. Rejection from those whose behaviors are substantially different will tend to strengthen in-group patterns of association and loyalty. Moreover, experimental research on opinion formation in small groups demonstrates that social interaction among group members tends to polarize groups by pushing members toward extremes [36, 111]. As a result, over time, cliques will tend to become more homogeneous internally, and more differentiated from other cliques in values and activities.

There are probably numerous traits on the basis of which individuals select their associates. Plausibly, criminal activity is one of them. In some circles, criminal behavior is strongly disapproved, and because it carries legal penalties and has the potential for incurring stigma, many individuals are likely to make it a criterion for friendship choice and partner selection for numerous activities—including those that do not involve crime. To avoid getting into trouble, or suffering the loss of reputation that might result from associating with troublemakers, committed law-abiders may avoid the company of those they believe to be engaging in illegal acts. For those who are more criminally inclined or who have less to lose from committing a crime (or being suspected of a crime), someone's reputation for criminality may be an attraction. It may connote glamor, masculinity, and defiance of conventionality. A reputation for involvement in crime may also discourage predatory violence, and be seen as extending protection to associates.

Discreteness might also be expected in the hypothetical case of criminal conduct being strongly determined by different alleles of a gene, or when potentially criminogenic social responses to different classes of people (whether defined by biological traits or social categories) are highly class-specific. In a society marked by strict gender distinctions that are rigorously enforced, one might find, for example, that the factors influencing male involvement in crime and delinquency are quite different from those that influence female involvement, or that they are the same but operate with different strength in each group.

On the other hand, there are considerations that might suggest a continuous distribution of involvement in crime. A number of the predictors of criminal behavior commonly employed in multivariate statistical analyses are continuous, or represent underlying psychological traits or propensities, or lifestyles (I.Q. scores, alienation from school, general strain, sensation-seeking, impulsivity and self-control, unsupervised leisure time) that we believe to be continuous. At the least this means that crime trajectories of individuals whose trajectories are assigned by finite mixture algorithms to the same group are unlikely to have exactly the same true trajectories. The same, obviously, holds for data sets representing aggregate rates or counts, when finite mixture modeling is used to study crime trajectories of neighborhoods [122] or cities [66]. Consequently, we do not expect the assumptions of the finite mixture model approach to hold exactly in typical criminological data sets. It is thus an empirical question how well the distributions of criminal histories are described by a set of discrete groups as compared with a continuum.

The great majority of researchers who have used finite mixture modeling to study criminal careers have not addressed this issue by determining the degree of heterogeneity within the groups that finite mixture modeling identifies. Moreover, defenders of the group-based approach [19], who point to the potential gains that might come from the use of finite mixture modeling, have not demonstrated that this approach is superior to others that might also be used to study criminal careers or other age-related outcomes. In commenting on the use of finite mixture models to approximate continuous distributions, Brame et al. [18] observe that “more research is required.” We second that observation, but extend it to the relative performance of different modeling techniques for the same data set.

In the present study, we explore the distributions of the parameters characterizing criminal career trajectories by employing two different strategies for statistical modeling that depart from the finite mixture modeling strategy by not imposing the assumption of discreteness on the trajectories—multilevel modeling and individual time series. Each has its strengths and weaknesses for this type of analysis. Carrying out the analysis with both models reduces the likelihood that the conclusions are artifacts produced by technical details of the estimation procedure. I conclude by offering advice to researchers contemplating the choice of estimation strategies.

Multilevel Modeling of Criminal Careers

The first of our two alternatives to the group-based approach entails the adaptation of multilevel modeling techniques to longitudinal data—in this case, the study of criminal events.

This approach treats the parameters that characterize criminal trajectories as continuous random variables, and thus permits the shape of their distributions to be assessed empirically, without a priori assumptions of discreteness or uniformity within classes [98].

However, as currently implemented in commercially available statistical software packages, the approach does assume that the parameters are drawn from a multivariate normal distribution, an assumption that may not hold exactly in all applications. After presenting our results, we will comment on the implications of this assumption, and then will present results using a different strategy that does not require the assumption of normalcy, namely individual time series.

For our analyses, we use a portion of a data set that has previously been analyzed using finite mixture modeling [30]. The data set contains information on the arrest histories of a large number of male youths who were confined by the California Youth Authority (CYA), and then tracked for a number of years following their release. We analyze data for two cohorts. The first was released from the CYA in 1981–1982; the second, in 1986–1987. Arrest records follow the subjects up to June 30, 2000, but are restricted to arrests made in the state of California. Regrettably, incompleteness of data make it impossible to adjust arrest counts for time in custody.

For this analysis I restrict our analysis to total arrests at each age, starting at age 7, without regard for the kind of offense (but excluding arrests for violating probation or parole conditions, or for traffic offenses). The 1981 cohort consists of 725 black, 503 Hispanic, and 761 “other” males; the 1986 cohort contains 532 black, 474 Hispanic,

and 437 “other” males. The great majority of those in the “other” category are white. Because the youths released in each cohort were of different ages at the time of their release, the maximum age at which they appear in the data set also varies from subject to subject. The minimum, mean, median, and maximum age of the first cohort are, respectively, 7, 22.07, 22, and 43; for the second cohort, they are 7, 19.77, 20, and 38. For more detailed information on the data, see Ezell and Cohen [30].

In the aggregate, arrests in the 1981 cohort peak at age 15, and then decline fairly smoothly. Arrests in the 1986 cohort also peak at 15, with a small secondary peak at about age 21. At age 28, they have declined to half the level they were at age 21. Finite mixture modeling of these data—specifically, the estimation of semi-parametric mixed Poisson models for the arrest counts—found optimal fits for six classes of offenders in each cohort, as assessed by the Bayesian Information Criterion [30]. Each of these classes has a trajectory with a distinct peak age, and a distinct rate at which involvement in crime declines after the peak.¹⁰ In the present analysis, I modeled the arrest trajectories in these cohorts using multilevel modeling techniques. I began by estimating a two-level model. The level-1 equation for individual i assumes that the individual’s arrest counts follow a Poisson distribution with a rate parameter λ_i that varies across individuals and whose logarithm depends linearly on age (measured in years) and the square of age. The rate parameter λ_i is thus assumed¹¹ to depend on age in this way:

$$\ln(\lambda_i) = \beta_{0i} + \beta_{1i}AGE_i + \beta_{2i}AGE_i^2 \quad (2)$$

Taking the logarithm of the rate parameter to be the latent dependent variable ensures that the expected value of the rate itself will always be positive. That would not be true if the rate parameter itself were to be modeled as a quadratic function of age. The postulated dependence allows us to identify curvilinearity in the age dependence of the logged arrest rate. This strategy for modeling criminal careers has long been used in criminal career research [42, 78, 100]. Based on the known aggregate age distribution of criminality [13, 31, 40, 42, 51], the coefficient of AGE is expected to be positive, and the coefficient of AGE^2 to be negative.

To allow for the possibility that the intercepts and slopes vary randomly across individuals, we specify a level-two model

$$\begin{aligned} \alpha_i &= \alpha + u_{0i} \\ \beta_{1i} &= b_1 + u_{1i} \\ \beta_{2i} &= b_2 + u_{2i} \end{aligned} \quad (3)$$

¹⁰ In Stata, I was able to extract eight groups based on the BIC criterion alone for the second cohort, but at least one of them had few members. The proportions Stata assigned to each group in the six-group model did not correspond exactly to those in Ezell and Cohen, who used Latent Gold in their work (2005).

¹¹ Some researchers add higher polynomials in AGE to the model, to capture asymmetry between the shape of the distribution below and above the maximum of the distribution. We restrict our models to a quadratic dependence on age. A longer follow-up period would probably call for a cubic term. Wang and Bodner [120] recommend restricting models to quadratic or lower powers in time, to avoid capitalizing on chance, except where theory or prior research provides a reason for including higher-order terms.

Each coefficient in the level-one equation is thus represented as the sum of a term that is the same for all individuals in the cohort, and a term that varies randomly across individuals. The random terms are taken to be multivariate-normally distributed with a mean of zero and a variance-covariance matrix

$$\tau = \text{Var}(u) = \text{Var} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{pmatrix} \quad (4)$$

I carried out all estimations in versions 5.1 and 6 of the HLM program using restricted maximum-likelihood, and allowing for over-dispersion in the Poisson regressions to accommodate the possibility of omitted sources of heterogeneity.¹²

In the first stage of the analysis, I estimated this model for all subjects in the two cohorts, and left the variables *AGE* and *AGE*² uncentered. To facilitate numerical computation, I divided *AGE* by 10 in doing the estimation, as recommended by Nagin [75, p. 44], but interpretations are based on the untransformed variable. Histograms of the empirical Bayesian estimates for the intercepts and the two slopes showed no evidence of discrete groups. All three distributions are unimodal, continuous distributions. At the same time, visual inspection of the histograms could not rule out the possibility that there are two or more highly overlapping distributions.

Because a substantial part of the research presented here relies on visual inspection of histograms, it requires mentioning that this method is imperfect. Imagine that a population is composed of two groups, and a variable being studied is normally distributed in both groups, with the same mean but with different variances. A visual inspection might find it hard to distinguish the histogram for this mixture from one in which a single group is present with a platykurtic distribution having the same mean. For criminological purposes, however, this inability to distinguish these two distributions might not be important.

It could also happen that in a multi-dimensional space, a distribution with more than one mode could, when histograms for each dimension or variable are considered individually, show just one mode on some of those dimensions. However, for a wide class of families of probability distributions, more than one mode will appear on at least one dimension. By looking at histograms for all three parameter estimates, we are somewhat protected against failing to detect a multi-modal distribution of cases.

When five dummy variables for the combinations of race-ethnicity and cohort were introduced into the level-2 model as predictors of the level-1 coefficients (with the dummy for the 1981-other category omitted as a reference), all proved to be statistically significant. On this basis, I conducted separate analyses for each of these three racial-ethnic categories within each cohort. Separate estimations for the different racial and ethnic categories allow for the possibility that in the aggregate, blacks, Hispanics, and others in the contemporary USA lead lives that are different enough in ways that are relevant for generating arrest histories to warrant separate analyses. Similarly, estimating separate models for each cohort enables us so to detect evidence for the historicity

¹² In a Poisson model, the variance of the distribution should equal its mean. Empirically, it is often the case that the variance exceeds the mean, a phenomenon called over-dispersion. It can have various causes, one of them being unmeasured heterogeneity [50].

of crime-generating social processes or criminal justice actions. By removing likely sources of trajectory clustering due to measured observables, this type of subsample analysis may also simplify the estimation of models by reducing the number of parameters to be estimated, and yielding subsamples that are more likely to be unimodal.

To test for individual variability in the parameters of Eq. (3), I first carried out a global significance test for the simultaneous vanishing of the diagonal elements in Eq. 4. for each of the six combinations of race-ethnicity and cohort. In each case, the test was significant, telling us that there exists some heterogeneity of trajectories. I then conducted tests for the significance of the individual diagonal elements in the variance-covariance matrix of Eq. 4, to determine just where the heterogeneity is located—the intercept, the slope of *AGE*, or the slope of *AGE*².

Table 6 shows the coefficients of the level-1 models for each combination of cohort and race-ethnicity. In addition to displaying the estimated coefficients, standard errors, and *t* tests, the table also shows reliabilities. These tell us how much of the variance in the data is accounted for by the variances of the slopes.

Table 6 Parameter estimates: multilevel models of total arrests, Poisson regressions

Fixed effects	Blacks 1981	Blacks 1986	Hispanics 1981	Hispanics 1986	Other 1981	Other 1986
Intercept	−4.275 (.1096) 39.008 .563	−5.244 (.1429) 36.690 .154	−4.616 (.136) 34.069 .222	−5.669 (.1624) 34.915 .540	−4.571 (.120) 38.169 .273	−5.664 (.1681) 33.691 .114
Age/10 slope	4.313 (.1028) 41.952 .655	5.776 (.153) 37.664 .189	4.697 (.137) 34.339 .267	6.211 (.1774) 35.017 .289	4.610 (.123) 37.346 .324	6.1175 (.1877) 32.586 .220
Age ² /100 slope	−1.0272 (.0246) 41.802 .645	−1.4839 (.0413) 35.911 .286	−1.147 (.035) 32.823 .367	−1.593 (.0480) 33.172 .374	−1.145 (.032) 35.690 .418	−1.5685 (.0517) 30.352 .336
Peak age	20.81	17.67	20.48	17.79	20.13	18.05
σ^2	1.774	1.829	1.835	1.898	1.855	1.887
No. of level-2 cases	725	532	503	474	761	437
Random effects-variance components (standard deviation)						
Intercept	1.6297 (.000)	1.769 (.000)	.6261 (.000)	1.442 (.083)	1.769 (.000)	1.215 (.064)
Age/10 slope	1.4273 (.000)	1.973 (.000)	.1616 (.000)	1.865 (.001)	1.973 (.000)	1.868 (.000)
Age ² /100 slope	.3656 (.000)	.578 (.000)	.0048 (.000)	.599 (.000)	.578 (.000)	.631 (.000)

Unit-specific model with robust standard errors. Entries for fixed effects are coefficients, standard errors (in parentheses), *t*-statistics and reliabilities; for random effects, they are the estimates and their associated probabilities

The age at which age of peak aggregate arrests is found by dividing the negative of the coefficient for *AGE* by twice the coefficient for *AGE*². We see, for example, that for blacks it dropped from 20.81 years old in the 1981 cohort to 17.67 years in the 1986 cohort. Only 2 of the 18 variance components fail to achieve statistical significance, indicating that there is significant variation in the intercepts and slopes across individuals. The two exceptions are the variance of the intercepts for Hispanics and for others in 1986. All of the components for the linear and quadratic slopes vary significantly across individuals. It is the shape of the distributions characterizing this variation that is of primary interest here.¹³ Our interest in the shape of these distributions lies in determining whether they are likely to have been drawn from relatively discrete underlying distributions. This enterprise calls for clarification of the concept of a group. Some natural phenomena come in sharply discrete categories. The elementary particles studied in physics, for example, can have spins that are integers or half-integers, and their charges can only be integral multiples of the charge of the electron (or in the case of quarks, 1/3 or 2/3 the charge of an electron). Intermediate values of spin and charge are not found in nature; both are inherently quantized. Discrete types or species also characterize patterns found in geology and biology and for some social phenomena like table utensils [70]. Social phenomena are not usually like that. Empirically observed distributions of interval-level variables typically studied in the social and behavioral sciences sometimes cluster, but imperfectly. They often appear to be drawn from continuous distributions that overlap to a certain degree. In these circumstances, taxonomic schemes that place objects into mutually exclusive groups will be most useful when they assign individuals to groups that are relatively homogeneous, and well separated from one another.

In social research, the researcher who is just beginning a study does not ordinarily know precisely what the shape of the distribution of an outcome variable is. In unusual cases, the researcher may be able to posit a shape on a theoretical basis, or, more commonly, to draw on previous work. Some researchers may ignore the issue and utilize an estimation method because it is familiar or because others writing on this topic have used it. Alternately, the researcher may try to infer the shape of the distribution by examining the empirically observed distribution in the data set being analyzed. That will be our strategy, and it will require a rule of recognition. How does one know whether, in a finite sample, more than one group is present?

For present purposes, we will infer that more than one group is present if the observed distribution appears to consist of random draws from a distribution with two or more modes that are well separated. Such distributions will look like overlays of two or more unimodal distributions with little or no overlap. There will be little variability within each group, so that most of the variability is between groups. To illustrate, when a one-way ANOVA is estimated for the means of the three-group model for the artificial beta density data analyzed previously, we find that the means are moderately well separated (.069, .209, .426) with low standard errors (.027, .056, .091). There is no overlap between the 95 % confidence limits for the means of the three groups. Then, 78.8 % of the total variance is between groups, 21.2 % within groups. This suggests that the groups are substantially but not perfectly homogeneous.

¹³ The intercept influences only the level of the curve, not its shape, which is governed by the linear and quadratic slopes.

If there is a great deal of overlap between groups, it might be the case that more than one group is present, but this information will be of limited value because there will be a great deal of uncertainty as to the group to which an individual trajectory should be assigned. In these circumstances, treating the empirical distribution as arising from one group might have the advantage of simplicity even though the fit could be improved marginally by positing more than one group. An additional useful screening criterion is that there be a non-trivial number of subjects in each group. This requirement will prevent us from overfitting a data set by treating each trivial departure from a smooth unimodal distribution as evidence of a distinct group.

In adopting these criteria, I am taking a social constructionist and pragmatic view of the question as to how many groups to use in classifying cases found in a data set. This view asserts that more than one group is present when it is likely to be helpful in research or practice to recognize their distinctiveness. Indeed, phrases like “well separated,” which I invoked above, are subjective. Any formal rule would be arbitrary, and at this point in the development of methods for studying criminal career trajectories, premature.¹⁴ The point at which it makes sense to say that two local modes in a distribution are well separated is likely to be context-specific. For the present, I prefer to leave the question an open one, rather than declaring a rule *ex cathedra*.

I supplement this somewhat subjective approach to group determination with a more formal approach by using Stata's *dipetest* for unimodality [24, 47, 48] to determine that a departure from unimodality is more than a random fluctuation from a unimodal distribution occasioned by the sampling procedure.¹⁵

For each of the six combinations of cohort and race-ethnicity, there are three distributions for the empirical Bayesian estimates—one for the intercept, one for the linear age coefficient, and one for the quadratic age coefficient. Limitations of space preclude the display of graphs showing all 18 distributions. To illustrate the results, I show the histograms for the estimated intercepts and slopes for black males in the 1981 cohort in Fig. 2a–c; and for Hispanic males from the 1981 cohort in Fig. 3a–c. Figures for the other four cohort-ethnicity categories may be had upon request.

The shape of the distribution of the intercepts for the 1981 black subjects can be seen from Fig. 2a. It is close to a normal distribution; for the 1986 cohort, which is not shown, it is somewhat skewed, with a concentration of intercepts on the low side. This signifies a reduction in the overall frequency of arrests. This reduction may reflect a change in the intake to the California Youth Authority—with the later cohort containing a higher proportion of youths with low rates of arrest. The distribution of the estimated coefficients for *AGE* is skewed, with the arrest rate for most of the subjects rising fairly rapidly as the subjects got older (see Fig. 2b). This skewness was stronger in 1986 than in 1981. Finally, the distribution of the estimated coefficients for *AGE*² is also skewed,

¹⁴ A recent attempt at greater precision can be found in Erosheva et al. [29], who propose that groups should be considered distinct “when there exists a time interval during which the 95 % credible bands for predicted individual trajectories show no overlap” (at p. 11.26). Why 95 %? Presumably, it is only because confidence intervals are, by convention, usually based on 95 %. Yet this convention has no substantive scientific basis. It is preserved as a customary procedure maintained mainly because there is no compelling reason for choosing a different figure.

¹⁵ This routine was written by Nicholas J. Cox, and is available for download from <http://fmwww.bc.edu/RePEc/bocode/d>. A test for multimodality has been developed by Henderson [49]. To the best of my knowledge, it is not yet part of commercially available statistical packages.

Fig. 2 **a** Histogram of estimated intercepts for blacks, 1981. **b** Histogram of estimated linear slopes for Age, 1981 black cohort. **c** Histogram of estimated slopes for age-squared for 1981 black cohort

with a concentration of cases on the low side (see Fig. 2c). This tendency, too, became more pronounced over time. There is no clear indication of distinct groups in any of these graphs. Each distribution is a fairly smooth, continuous, unimodal distribution with a single peak. Were there separate groups containing substantial numbers of boys, we would be seeing more than a single peak in each graph.

Parallel estimations for the Hispanics in the same cohort find intercepts that are somewhat lower than for blacks, pointing to a somewhat lower frequency of arrests. The age at peak arrest in each cohort is only slightly different from that of blacks, but is almost 3 years younger in the 1986 cohort than in the 1981 cohort. The distributions of the estimated intercepts and coefficients of age and age-squared again show no evidence of sharply defined discrete groups. The shape of the distributions for Hispanics in the 1981 cohort is, however, somewhat more irregular than the corresponding distribution for blacks in that cohort. The distributions of intercepts and slopes look like the superposition of a narrow peaked distribution toward the high end, on top of a broader, less peaked distribution at the low end, creating a continuous but bimodal distribution. The linear and quadratic slopes are skewed and somewhat irregular.

Fig. 3 **a** Histogram of estimated intercepts, 1981 Hispanic cohort. **b** Histogram of estimated linear slopes for age, 1981 Hispanic cohort. **c** Histogram of estimated quadratic slopes for age, 1981 Hispanic cohort

The histograms of estimated coefficients for the other four groups—blacks and Hispanics for 1986, and others for 1981 and 1986, display essentially the same patterns. Either the distributions are smooth, moderately skewed and unimodal, with no indication of multiple peaks, or there are two very substantially overlapping distributions (as seen, for example, in the histogram for “others” released in 1981). None of the histograms shows evidence of anywhere near the six distinct groups identified through finite mixture modeling as characterizing the cohort as a whole.

As indicated above, I supplement these graphical procedures with a formal statistical test for unimodality by using the diptest. Each test draws 10,000 random samples from a uniform distribution, and uses the distribution of estimates as the basis for the test. By using the uniform distribution for the comparison with the observed distribution, this procedure provides an upper bound to the departure of the distribution from unimodality. The test results are found in Table 7.

Only for the 1981 cohort of Hispanics are the significance tests statistically significant at the .05 level. Because (a) the diptest uses an extreme case for purposes of comparison, (b) because the diptest does not show evidence of more than one group in the 1986 cohort of Hispanics, (c) because the two groups that appear visually in the histogram overlap very strongly if they are present, and (d) the independent time series estimations we report below do not show indications of two or more groups for the

Table 7 Diptests for unimodality (based on empirical Bayesian estimates)

Cohort-Ethnicity	Coefficient	Dip	<i>p</i>
1981 blacks	Intercept	.0089	.9884
	Linear slope	.0095	.9746
	Quadratic slope	.0092	.9843
1981 Hispanics	Intercept	.0389	.0040
	Linear slope	.0434	.0005
	Quadratic slope	.0240	.1471
1981 others	Intercept	.0526	.5465
	Linear slope	.0507	.4113
	Quadratic slope	.0730	.1762
1986 blacks	Intercept	.0180	.0883
	Linear slope	.0064	.9976
	Quadratic slope	.0054	.9990
1986 Hispanics	Intercept	.0130	.4584
	Linear slope	.0088	.9506
	Quadratic slope	.0071	.9978
1986 others	Intercept	.0280	.8058
	Linear slope	.0356	.4562
	Quadratic slope	.0251	.9105

parameter estimates for this cohort, these results do not point to a major departure from the generalization that all six of the cohort-ethnicity groups have trajectories that are well described by smooth, somewhat skewed unimodal distributions for the intercepts, linear, and quadratic slopes characterizing the arrest histories. If there are distinct classes of trajectories present in this data set, they do not differ from one enough for criminological researchers to care much about them.

Before drawing out the implications of these findings for criminal career analyses, we consider a complication associated with the hierarchical linear modeling approach. The empirical Bayesian estimates of coefficients are biased. They are computed as a weighted average of a coefficient generated for each subject from that subject's own arrest history, and a coefficient generated by the group as a whole. The weights vary with the amount of information available from each subject. The fuller the information available on a given subject, the more heavily that subject is weighted in the computation of the empirical Bayesian estimate [98]. Moreover, the assumption of multivariate normality made by multilevel modeling routines may not hold exactly in the data being analyzed. Consequently, we must consider the possibility that a failure to find stronger evidence for discrete groups is due to bias that leads the method to underestimate or misrepresent the variability in the estimates. But in a sample as large as ours, and with as many observations per subject as are present here, one would not expect this bias to be large enough to reduce a distribution with two or more prominent modes to a unimodal distribution, at least if there are reasonable numbers of cases in each group.

To confirm that this is the case, I created an artificial data set with two groups, and then estimated hierarchical models that do not assume the existence of discrete groups. Specifically, the artificial data set has 25 annual observations for each of 100 individuals,

between ages 15 and 39. These ages are comparable to those in the Ezell data set, and to those in other data sets used to study criminal careers. Note that this is a sample substantially smaller than in the actual Ezell data set, which means that it has worse statistical power—that is, the ability to detect differences if they are really present. Fifty of these individuals were given time-dependent Poisson rate parameters defined by the equation

$$\lambda_1 = \exp(2 + .1t - .005t^2) \quad (5)$$

The other 50 individuals have Poisson rate parameters defined by

$$\lambda_2 = \exp(.05t - .001t^2) \quad (6)$$

Both sets of individuals have trajectories of the same general shape—first rising with increasing time, and then declining, but with different peak ages. I then created a dependent variable whose values are random draws from a Poisson distribution with rate parameters defined by the above two equations. By construction, the data set is perfectly suited to the finite mixture approach, as it was created with two distinct groups, each having its own trajectory, and with all the individuals in each group having exactly the same trajectory parameters. Individual differences in criminal careers are thus due only to group membership, and to the randomness of the Poisson process that turns a trajectory into a sequence of crimes.

I then used version 5.1 of the HLM program to estimate a mixed-effect Poisson regression model on the entire data set, specifying the intercept and two slope coefficients as random variables. The program was unable to estimate the model, and suggested that one of the random coefficients be respecified as fixed. I therefore chose the coefficient of the quadratic term to be fixed (i.e., not varying randomly across individuals) and repeated the estimation.¹⁶

This model could be estimated without difficulty. Histograms for the individual estimates of the intercept and slope coefficients are displayed in Fig. 4a, b. By inspection, there are two distinct groups; the groups are well separated, with no signs of overlap.

As for the concern about the possible departure of the random contributions from a normal distribution, several investigators have found that a misspecification of the distribution of random effects does little damage to the estimates of the fixed effects so long as there are many observations. As that number increases without limit, the estimates converge to their true values regardless of the distribution of random effects [22, 67, 86, 117]. In finite samples, of course, violations of the distributional assumptions for the distribution of random effects can be a source of bias [110]; however, the estimates of the fixed effects are only slightly affected. There is bias in the standard error estimation, but it can be corrected [22, 116, 117].

¹⁶ This is a common practice in this circumstance [see, for example, 33]. I was able to estimate the model with all coefficients treated as random in version 10 of Stata. The existence of two groups was clearly evident. This model took many hours to converge. The estimates of the fixed parameters were not close to their true values, and the distributions for the constant term and linear term in *age* appeared to be unimodal. These results suggest that there may be estimation difficulties when groups are present but not distinct enough to be distinguished readily in modestly-sized samples.

Fig. 4 **a** Histogram of estimated intercepts, two-group artificial data. **b** Histogram of estimated slopes, two-group artificial data

In principle, concerns about potential bias stemming from departures from the normal distribution in the random effects could be remedied. There is nothing inherent in multilevel modeling to preclude the estimation of models with error terms whose distributions are other than normal. Statisticians have proposed more flexible alternatives, such as Student's t or a skew-normal distribution [3, 23, 68]. Tong and Zhanag [114]. To the best of my knowledge, software allowing for these options is not commercially available.

To determine the accuracy with which individuals are being assigned to the appropriate group by this procedure, I dichotomized the individual estimates of the intercept and linear slope at the mean value of each distribution, and then cross-tabulated these estimates with the group assignment. This cross-tabulation showed perfect success in cross-classification. Each individual with a high true intercept or with a high slope was assigned to the group with high estimates of the intercept or slope.

A full assessment of the capacity of multilevel modeling to separate out distinct groups when they are really present in the data lies beyond the scope of this paper. Nagin [75] observes that there may be some complex patterns of trajectories that could not be distinguished easily in a multilevel analysis, but that would be distinguished in a group-based approach. This may well be the case. Whether there are empirical examples of such patterns in the data sets criminologists typically analyze when studying criminal careers is another question, one that has not been seriously examined thus far.

Most researchers who analyze criminal careers have used only one method to analyze their data, and so the possibility of comparing the results of the two approaches on the same data set has been little explored in empirical research on criminal careers. I did recreate the artificial data set that Nagin [75] used to illustrate the group-based approach; like our artificial data set, it consists of two sets of differently specified trajectories. His analysis shows that the SAS routine PROC TRAJ recovers their shapes well. In this instance, that there are two distinct patterns of trajectories is evident on inspection of a scatterplot of the empirical trajectories. In this case, such an inspection would show a researcher right at the outset that a finite mixture modeling approach would be appropriate. In any event, the results of this preliminary investigation suggest that the multilevel approach need not be avoided for fear that the approach itself is incapable of discerning genuine group differences.

Individual Time Series

To ensure that the conclusions reached in the previous section are not a result of the empirical Bayesian estimation or the assumption of multivariate normality for the errors in the level-2 equation, I repeated the analysis using individual time series (ITS). The ITS approach estimates separate Poisson regressions for each subject in the data set. It saves the estimated intercepts and slopes for each estimation, and then examines their empirical distributions. This approach avoids the bias produced by an empirical Bayesian estimation, and makes no assumptions about the distribution of error terms in a level-2 model. This procedure has previously been used by Bushway et al. [21]. I carried out this procedure for each of the six race-ethnicity-cohort combinations using the *statsby* command in version 11 of Stata.

Figure 5a–c shows histograms for the estimated intercepts, linear and quadratic slopes for the cohort of blacks released from custody in 1981; Fig. 6a–c shows the same histograms for the cohort of Hispanics released in that year, but they have been truncated. The reason for the truncation is that all the histograms based on the complete data set show a small number of cases whose intercepts and linear slopes are much lower than those of the bulk of the cases, and whose quadratic slopes are much higher. These outliers are cases

Fig. 5 **a** Histogram of estimated intercepts (ITS) for the 1981 black cohort. **b** Histogram of estimated linear slopes for the 1981 black cohort (ITS estimates). The slopes are for $(age/10)$. **c** Histogram of estimated quadratic slopes for the 1981 black cohort (ITS estimates). The slopes are for $(age/10)_2$

Fig. 6 **a** Histogram of estimated intercepts (ITS) for the 1981 Hispanic cohort (ITS estimates). **b** Histogram of estimated linear slopes for the 1981 Hispanic cohort (ITS Estimates). **c** Histogram of estimated quadratic slopes for the 1981 Hispanic cohort (ITS estimates)

for which there is very little information about arrests. In other words, these subjects had hardly any arrests in the entire follow-up period. In a multilevel model, their parameter estimates would be pushed toward the center of the distribution by the Bayesian averaging that does not take place when estimating individual time series.

In some research, these extremely low-level offenders might be of particular interest, and could legitimately be singled out for special attention regardless of whether they represent the extreme end of a continuous distribution or a separate class with very few members. However, if interest lies in the patterns that typify the bulk of the population, the multilevel approach, which obscures their existence, may be preferable. When these outliers in the extreme tail ends of the distributions are included in the histograms, the range represented by the horizontal axis is widened to the point where the great majority of the cases are squashed together in a manner that makes it impossible to see the shape of the distribution in detail for the great majority of the cases. By dropping the outliers, we are in a better position to see meaningful structure if it is present for the most frequently arrested offenders.¹⁷

In most respects, the histograms confirm the conclusions reached using multilevel modeling. The distributions show smooth, unimodal distributions, along with a very

¹⁷ Alternatively, a zero-inflated Poisson or negative binomial regression could be estimated to accommodate these cases in a single analysis. Because there are so few cases exhibiting this arrest pattern, we elected not to go that route.

small number of cases whose intercepts or slopes are much lower than those of the bulk of the cases. Significantly, this is as true for the Hispanics in the 1981 cohort as it is for the other five groups. Their histogram does not display the appearance of bimodality observed for that cohort in the multilevel modeling analysis.

Diptests for the distribution of the estimates obtained from the individual time series can be found in Table 8. In all but one of the 18 tests, the test statistic is not statistically significant at the .05 level. The one significant test is for the intercepts in the 1986 cohort of blacks. However, the truncated histogram for those intercepts shown in Fig. 7 shows no indication of bimodality. Taking the histograms and diptest results together, we see no strong evidence that there are distinct classes of arrest histories among the alumni of the California Youth Authority.

Implications for Theory Testing

Several commentaries on the group-based approach to the investigation of criminal careers have observed that many published studies can be considered exploratory [19, 91, 120]. They do not begin with a theory specifying a number of groups or the variables that determine whether a given individual's criminal career follows one trajectory rather than another. Rather, they estimate models with different numbers of groups, adopt the best-fitting model, graph its trajectories, and ascertain how well covariates predict group membership.

The use of the group-based method in exploratory research is unobjectionable. Exploratory work can be important in fields where theory is lacking. Where theory

Table 8 Diptests for unimodality (based on individual time series)

Cohort-Ethnicity	Coefficient	Dip	<i>p</i>
1981 blacks	Intercept	.0092	.9847
	Linear slope	.0098	.9679
	Quadratic slope	.0017	.9985
1981 Hispanics	Intercept	.0111	.9836
	Linear slope	.0097	.9869
	Quadratic slope	.0128	.9235
1981 others	Intercept	.0105	.9351
	Linear slope	.0078	.9976
	Quadratic slope	.0238	.0599
1986 blacks	Intercept	.0308	.0290
	linear slope	.0079	.9985
	Quadratic slope	.0099	.9935
1986 Hispanics	Intercept	.0129	.9390
	Linear slope	.0093	.9969
	Quadratic slope	.0107	.9929
1986 others	Constant	.0125	.9677
	Linear slope	.0119	.9833
	Quadratic slope	.0088	.9975

Fig. 7 Histogram of estimated intercepts for the 1986 black cohort (ITS estimates)

exists, however, researchers may want to move beyond exploration to test theories using the group-based approach. In a thoughtful discussion, Brame et al. [19] point out that testing is important to the development of theoretically based, empirically validated science. When a theory's predictions are disconfirmed by empirical evidence, the theory should be abandoned or modified. They encourage finite mixture modelers to go beyond description by testing theoretical predictions as to how many groups are present in a data set. Yet our findings suggest that inferences about the existence of distinct latent classes must be made with great caution when using the group-based approach. Where a theory predicts a specific number of groups, confirming that this number emerges as optimal for a data set may not mean too much because of the method's tendency to identify too many groups when the empirical distributions differ from those assumed to characterize the data-generating process. This could lead to an inaccurate acceptance of a theory, or to its erroneous rejection. For this reason, we question the conclusion of Nagin and Odgers [79] that growth-based trajectory modeling "is ideally suited for analyzing the influential taxonomic theories of antisocial and delinquent behavior of Moffitt [71] and Patterson et al. [88], [89]." When a theoretical prediction about the number of groups is being tested, it is imperative to use a method that can be counted on to yield valid conclusions as to the true number of groups present in a data set.

The same considerations apply when testing other criminological ideas about the processes at work in the unfolding of criminal careers—for example, looking for evidence of state dependence—in which past criminality has an ongoing effect on future criminality [30, 80], or for assessing the impact of exogenous changes in the level of strain or social control. If the importance of state dependence or strain is greater for some research subjects than others, it would be important for researchers to get the number of trajectory classes right.

In testing a theory, it would also be important to rule out alternative theories that might make the same predictions regarding the number of groups. Etiological theories in criminology specify the processes believed to bring about criminal behavior, or abstention from criminal behavior. Consequently, even if a

researcher felt confident that the estimator was yielding the right number of groups, the researcher would still want to see whether the theory accurately predicted the shape of the trajectories [5], and that the predictors of group membership are as predicted theoretically.

Occasions where researchers want to test predictions regarding the number of groups might not be many. Few theories predict a precise number of discrete groups. Etiological theories of criminal activity—at least those originating in sociology—are usually general theories involving the effects of continuous variables on outcomes. Whatever groups are present in the data are products of the distributional patterns of the predictors of criminal involvement (social disorganization, general strain, inept parenting, and so forth). The explanation of how these causes of crime are distributed in a population generally lies outside the scope of criminology. The distribution of these causes, moreover, is likely to vary from one time and place to another, which means that studies of different populations are likely to differ as to the number of optimal groups delineated by group-based fitting procedures, and the proportions of individuals following distinct trajectory patterns.

This variability in the distribution of causes could arise not only in relation to the study of social causes but also in studying genetic causes, because the distribution of alleles for genes relevant to criminal behavior could vary from one time and place to another. For these reasons, Nagin and Odgers [79] are right to point out that “one should not expect the numbers and shapes of trajectory groups to remain constant across samples from different populations.” The variability in the number of groups found optimal in different empirical studies [106] provides empirical confirmation of this observation. The number is clearly not a law of nature. For this reason, it may be a misdirection of research energies to consider the determination of the right number of groups to be the primary goal of research (though it has a legitimate place in the process of studying the factors that influence the shape of criminal careers).

Some further implications of the present study for theory testing may be noted. Traditional variable-oriented etiological research on crime has focused primarily on factors that influence levels of involvement. In trajectory analysis, once research has identified an optimal number of groups for the trajectories, attention shifts to the identification of factors that predict membership in the most probable trajectory group, or that influence the probabilities of belonging to each group. These estimates are conditional on the number of groups that characterize the model adopted. There is, however, some uncertainty associated with model selection that will have implications for the certainty with which researchers can assess the influences of individual traits on the probability that an individual has a particular trajectory [12, 25, 99]. The available software allows for conventional inferential statistics like *t* tests to be used in this assessment, seemingly without adjustment for the uncertainties associated with the choice of number of groups, or with decisions as to the order of the polynomial in time or age associated with each trajectory. If, as the research presented here suggests, the identification of distinct trajectories is problematic, this important stage in the research could also produce estimates whose interpretation is unclear.

In addition, the group-based approach may identify trajectory classes that are not well distinguished from one another, or separated by level-2 variables. In commenting on a new analysis of the Cambridge Study in Delinquent Development, Macmillan [65] observes that the five trajectories of convictions identified by the researchers [95] using

the group-based method all tend to rise, peak at around age 18, and then decline. Here, the payoff from using the high-powered machinery of finite mixture modeling may be low. To take another example, Ezell and Cohen [30, p. 177] observe in their study of the California Youth Authority releases that “there are few variables that have any numerically distinguishing values across the latent classes.” Where this is so, the group-based approach may tell us only that some of those released desist from crime faster than others, without telling us what causes the difference. In other data sets, of course, the level-2 predictors may have greater explanatory power. Proponents of finite mixture modeling have touted the ability of some group-based research to identify traits that successfully distinguish the trajectories. For purposes of testing etiological theories of criminality, this ability is critical. What is not yet clear is that they do so more successfully than other estimation strategies, such as multilevel modeling.

Some other considerations relevant to theory testing in trajectory analysis that are not the focus of the present study have to do with limitations in current statistical theory and software. In recent decades, causal analyses of many social phenomena have identified potential issues that need to be addressed in causal inference with longitudinal data. One has to do with the nesting of individuals in locations. Among other things, this leads to non-independence of observations, creating a problem for inferential statistics. Multilevel models are designed to handle this lack of independence. Non-independence of observations is potentially relevant for finite group modeling as well, because the statistical theory underling Bayesian model selection procedures also rests on the assumption that observations are independent. It is not clear how a trajectory analysis might take the non-independence of spatially distributed data into account.

Another issue is endogeneity. One source of endogeneity in a set of variables is reverse causation. Another is omitted variables that are correlated with predictors. Instrumental variable methods have been developed to handle these circumstances. Fixed effects methods of estimation for panel data have been developed to purge estimates from bias due to the omission from a model of time-invariant variables [17]. This is a potential issue when using multilevel methods as well [93]. Causal analyses may reach biased conclusions if they fail to deal properly with omitted variable bias [90]. The group-based method deals with unmeasured stable sources of heterogeneity differently, and in a manner that may not be amenable to the modeling of reverse causality. A multilevel framework, on the other hand, should be able to handle reverse causation readily. We have not, however, encountered discussions of how to address them when doing finite mixture modeling.

A third consideration has to do with nonstationarity. Time series and panel data econometrics of the past three decades have shown that when longitudinal data are not stationary, spurious regressions are possible, and conventional inferential statistics can be biased. To the best of our knowledge, the statisticians who have developed multilevel modeling and finite mixture modeling have not considered how to incorporate these insights into their modeling strategies.

While these remarks have focused on the testing of theory, it should not be forgotten that empirical findings obtained in exploratory empirical research can stimulate theorizing to explain them. D’Unger et al. [27], for example, argue that an advantage of the group-based approach is that it can enable researchers to discover latent classes “where only a single observed class had been previously assumed and, therefore, theorized and empirically studied.” This passage suggests that an empirical study using the group-

based method can reveal the existence of a structure that can provide grist for the theoretical mill. In particular, the discovery of multiple, seemingly distinct groups with different trajectories can direct theorists to consider influences on criminality that display clustering or distinct classes, while their absence might encourage exploration of factors that vary smoothly. I agree. However, the inferential process could be thrown off if the estimation yields to a misleading conclusion as to the optimal number of groups. For these inferences to work well, it is important that the exploratory methods chosen not create artifactual groups. Otherwise, the theorist could end up pursuing a phantasm.

Future Research

On the basis of the analyses presented here, for male youths released from the California Youth Authority in the 1980s, arrest trajectories do not group into identifiable, statistically and substantively meaningful discrete classes. That finite mixture modeling yields too many groups when used to analyze data from the California Youth Authority cohorts suggests the desirability of re-analyzing data that have been studied previously using the group-based approach. We suspect that with these data sets, too, a re-analysis will show that there are fewer groups than have been found by previous researchers. Of particular interest in these re-analyses will be the question of whether conclusions about the effects of subject traits on trajectory parameters need to be modified.

Further empirical research will be needed to determine the distribution of trajectory parameters in other populations. The CYA data set consists of youths who at the start of the study had already been committed to a juvenile institution for an offense. In a study of individuals not selected for previous criminality, it is possible that the results could be quite different. In the population at large, for example, a substantial number of youths might show no criminality or very little, and could continue in this way for long periods of time. They might be a relatively discrete population, with a modest number of “late starters” taking up active criminal activity in response to some precipitating life event in adulthood. At the same time, those who are more actively criminal when young might be characterized by the kinds of continuous distributions seen here. Indeed, in a re-analysis of criminal career data for the London cohort of working-class males previously studied by Farrington [32], using the methods presented here, concludes that there appear to be two groups in the data (apart from non-offenders), not the four found by previous researchers using finite mixture modeling [27, 78]. This is a situation for which zero-inflated models might be appropriate, whether they are estimated through multilevel modeling or finite mixture modeling.

Implications for Researchers

Our findings have implications for researchers studying criminal careers. When researchers are interested in the shape of the distributions that characterize criminal activity in some population, or in identifying the factors that influence the shape of the trajectories that make up criminal careers, they should avoid assuming at the outset that criminal careers come in discrete clusters. It would seem preferable to begin by

recognizing that trajectories being modeled could be continuous, and look for clustering of trajectory parameters estimated under that assumption. If there is no evidence for the existence of discrete groups in a multilevel analysis, then the research might proceed using methods based on continuous distributions.

Bushway et al. [21], whose goals are somewhat different from ours, conclude in an analysis of Dutch criminal careers that the multilevel modeling approach was superior to both finite mixture modeling approach with seven groups and individual time series analyses in providing a parsimonious and accurate model of individual trajectories. Their findings and those presented here suggest that my recommendations may be useful in the analysis of criminal careers and developmental patterns in other data sets. The simulations carried out by Sterba et al. [110] conclude that sometimes finite mixture modeling leads to more precise estimates of model parameters, and sometimes multilevel modeling does. Some of the features of a data set that make one method outperform the other will not usually be known to the researcher (e.g., the shape of the distribution of the latent random effects). For this reason, the researcher might be well-advised to consider both approaches. Their conclusion that the multilevel estimates of slopes tends to outperform the estimates obtained by the group-based method in the presence of skewing reinforces our recommendation that the multilevel approach should not be overlooked in analyzing criminal career data.

On the basis of the research presented here, the qualitative examination of histograms for the distributions of trajectory parameters estimated through multilevel modeling seems to be a potentially informative way to begin a study of criminal trajectories. However, when comparing models involving groups with those that assume continuous distributions, researchers can, if they wish to use formal goodness-of-fit statistics to aid in the comparison, use the BIC statistics in the comparison. To date, trajectory modelers have used this method only to determine the optimal number of groups use in finite mixture modeling. But there is nothing to prevent researchers from comparing the BIC for the best-fitting group-based model with the BIC from the best-fitting multilevel model.¹⁸

To illustrate this procedure using our artificial data for beta regression, I estimated several different models, beginning with an OLS regression of the outcome with no predictors and no level-2 variance (there are no level-2 variables in the data set). This model yields a BIC statistic of -1084.971 . A finite mixture model with two components designated as Poisson distributions yielded a much higher BIC statistic of 1030.371 , with 100 % of the cases in group 1 and 6.21×10^{-9} in the second group. These estimates are clearly telling us that there is no second group to consider. In this case, at least, the skewness assumed for the individual components prevents the estimation process from finding too many groups. Finally, a fit using Stata's *betafit* command, which carries out a beta regression, yielded a BIC value of -1231.787 . A one-group model assumed to have a gamma distribution had a worse fit (BIC =

¹⁸ In addition, comparison of BIC statistics could help researchers determine the nature of the components being assumed by a model. In our artificial example, the two-group mixture of gammas has a lower BIC statistic than our three-group mixture of normal distributions. Most of the diagnostic criteria are well-satisfied. However, the two trajectories are poorly separated, providing a clue that a multigroup model may be inappropriate.

–1202.358). Models with two or more groups and that incorporated other assumptions about the distributions (such as lognormal or Student's t) did worse. The Bayesian factor comparing this with the Poisson model identifies the beta distribution as the most appropriate.

That the *betafit* model should yield the smallest value of the BIC statistic in this data set is not surprising.¹⁹ The variable being analyzed was constructed to have a beta function as its distribution. In analyzing real data, one would ordinarily not know the true distribution, but the range of values and its empirical shape revealed by a histogram should be able to point the researcher to the types of functions that should be considered. Users of some of the most widely used software for conducting group-based analyses, however, may have a limited range of models from which to choose.

To illustrate this procedure with real data, we return to the Ezell and Cohen data set, and compare a multilevel analysis with a finite mixture model for the 1981 cohort of blacks. Because the outcome is a count of arrests, I estimated Poisson and negative binomial regressions. The best mixed model for the negative binomial regression has a BIC statistic of 49647.68; for a three-group Poisson model, it is 51380.04. The four-group Poisson model did not converge. The diagnostic criteria recommended by Nagin [75] all look good for the three-group model. The multilevel models for the 1986 cohorts of blacks and of whites and others also fit the trajectories better than the corresponding finite mixture models. For the other three cohorts, it was the reverse.

Researchers wedded to the group-based method and who did not consider other modeling strategies would choose suboptimal models by failing to consider a potential alternative. That each approach provides better fits for half of the six cohorts makes it hard to believe that the model selection criteria are necessarily telling us something fundamental about the processes governing the shape of the arrest trajectories. The extension of the model selection procedure currently being used by finite group modelers thus confirms the conclusion reached through the analysis of histograms that there is no strong evidence for discrete groups in this particular population.

For researchers concerned only with optimal curve-fitting, this is not upsetting. For these researchers, the important “take away” message is that consideration of both types of models has the potential for obtaining an improved fit. For someone determining the “true number of groups,” this comparison of approaches suggests that the procedures now commonly being used in group-based analyses to determine the number of groups for a model are not to be trusted (unless supported by other evidence), and should not be used for this purpose, even where they provide a better fit than multilevel models.

With results from both approaches, we can also ask whether we would come to substantively different conclusions about the nature of the trajectories being modeled. For the 1981 cohort of blacks, the trajectories modeled with three groups peak at 20.56, 22.04, and 21.31 years. In the two-group models, they peak at 21.30 and 21.36 years. This compares with the peak at 21.35 years found with the negative binomial regression that made no assumptions about discrete groups. These results make clear why the histograms for the estimated coefficients from the multilevel analysis did not appear as separate modes: they are so similar that the histograms would not be capable of distinguishing them. With trajectories for the groups having maxima that are so similar,

¹⁹ Stata's *betafit* routine also performed well in recovering the model parameters. Its estimates of the model parameters were 2.06 for alpha and 6.33 for beta.

one would be hard put to argue that the group-based approach yields new insights into criminal careers by allowing each group to have its own trajectory. In studies of other populations, the group-based approach might yield more substantive findings.

Thus far, we have been assuming that the purpose of the research is to investigate the nature of heterogeneity in criminal careers. This means describing the empirical patterns of crime or arrest trajectories, and then explaining any differences that may be observed. Many developers and uses of the group-based approach have observed that finite mixture modelers need not assume that the world is actually divided into discrete, non-overlapping groups; this approach can also be used to approximate a continuous distribution for purposes of describing a population. Nagin [75], a leading contributor to the development of group-based methods, has opined that the method be seen as providing a simple, easily grasped representation of a more complex reality. Many users have adopted this understanding of the procedure. This paper picks no quarrel with it. However, for purposes of finding out as well as possible what that distribution is—and in particular, whether it has discrete groups, no *a priori* advantage is to be had in approximating that continuous distribution with a discrete one, given that statistical methods for analyzing continuous distributions are also available. The existence of discrete groups should be a conclusion of research (when analysis points in that direction), rather than a premise adopted at the start.

An analogy to OLS regression may help to clarify the view of the role that the group-based method should play in descriptive and explanatory research being advanced here. Consider a hypothetical data set consisting of observations on two interval-level variables, x and y , in which x causes y in a manner that conforms to all of the assumptions of OLS regression. Clearly, in this circumstance, OLS regression would be the preferred way to analyze the influence of x on y ; it is the best linear unbiased estimator. But there is nothing to prevent us from aggregating the observations of y into a finite number of discrete ordered categories, and carrying out the estimation using ordinal logistic regression. It would not be wrong, but it would pointlessly throw away information without achieving any gain, and would entail the adoption of a model that is unnecessarily complex. One could justify the procedure by saying that the groups created in this way are just an approximation of a more complex underlying reality, but it would still be the case that this would not be the optimal way of analyzing the data. On the basis of the results presented here, the use of the group-based approach to the study of careers may be comparably suboptimal in some of the studies that have used it.

This logic, then, implies no objection to researchers exploring their data set with group-based approaches. It can be anticipated that such analyses will “reveal” the presence of multiple groups, as has been the case with every study of criminal careers modeled as finite mixtures to date. This is likely to be true because numerous studies, starting with that of Wolfgang et al. [124], have found that in the general population, the level of involvement in crime is highly skewed. Most individuals commit few offenses per year; a minority commit more. This will produce a skewed distribution for the intercepts.

My results lead me to recommend that researchers undertake a comparison with a multilevel analysis before concluding that there are in fact discrete clusters of criminal career trajectories. The researcher who tries both methods and finds discrepancies as to the number of groups may want to examine the distribution of parameters estimated by

the multilevel modeling approach. If they are substantially skewed, this should raise a suspicion that the finite mixture approach is finding too many groups due to the departure from normality. In addition, the researcher might estimate separate multilevel models for the subjects assigned by a finite mixture approach to each particular group. If there is substantial variation in trajectories within each group, this might also raise questions as to whether the finite mixture approach is optimal.

Erosheva et al. [29] observe that publications using the group-based method usually do not examine heterogeneity in trajectories for subjects classified as having a particular trajectory. Moreover, they point out that graphs showing only the theoretical trajectories for an optimal fitting model may lead readers to think such within-group heterogeneity is minimal, even when it is substantial. As we saw above, an ANOVA could be informative here. The models discussed by Kreuter and Muthén [58] offer an appealing way to introduce heterogeneity within trajectory groups where evidence of its existence is found.

As for the contention that finite mixture modeling is simpler, this will not always be the case. One way of measuring the complexity of a model is to count the number of free parameters that must be estimated. When fitting a quadratic equation in time in a group-based approach, there will be three parameters for each group. In addition, there will be estimates of the proportions of individuals assigned to each group. Thus, if the optimum fit is provided by k groups, there will be $3k + (k-1)$ parameters. If the model being estimated posits the existence of 5 groups, 19 parameters will be fit.²⁰ If there really are five discrete groups, this number may be necessary to capture the full complexity of the data. However, if the trajectories are really continuous, the group-based model will lead to a model of unnecessary and misleading complexity.²¹ This complexity also creates problems for numerical estimation [120].

Where the results of a multilevel model point to the existence of relatively discrete groups, the researcher will ordinarily want to understand their origins. This investigation can begin by introducing likely predictors of group membership into a multilevel model as level-two predictors, provided that data for these predictors exist in the data set. I looked at the effects of race-ethnicity in just this way. Alternately, the researcher could estimate a finite mixture model by assigning to each individual that individual's most likely trajectory, and then use the predictors of group membership to predict each individual's trajectory in a logistic regression. In a variant of this procedure utilized by Bushway, Sweeten, and Nieuwbeerta, one could weight each of the individual trajectories by the probability of the individual having that trajectory.

In data sets that track subjects over long periods of time, and that contain many observations per subject, individual time series should also be estimated, as a way of obtaining information about the distribution of parameters that does not depend on any assumptions about their distribution. In short time series, there will be too few observations to obtain precise estimates of the parameters used to model the trajectories in this way. A researcher who wanted to go beyond the eye-balling strategy for

²⁰ Some researchers add a cubic term in *AGE* to our Eq. (2) to accommodate asymmetry in the shape of the curve. If this is done, then there will be $4k + (k-1)$ parameters to estimate. When 5 groups are estimated, this comes to 24 parameters.

²¹ In Bushway et al.'s [21] comparison of modeling approaches, the best-fitting group-based model, with 7 groups, had 34 free parameters, compared to 14 in the latent growth curve model fit to the same data. The individual time series had 18,460.

assessing whether the histograms of parameter estimates for the individual trajectories estimated through our ITS strategy could look for evidence of grouping by carrying out cluster analyses of the trajectories. Many options for cluster analysis are available in commercial statistical packages.

In choosing how to proceed, researchers may want to consider that the multilevel approach has some advantages over finite mixture modeling for those data sets in which there are no sharply delineated distinct groups. The multilevel approach allows for the possibility that the factors predictive of intercepts (levels of individual criminality) could differ from the factors that influence the linear and quadratic slopes in the level-one equation. These factors govern the age at which individuals reach their peak involvement in crime, and the speed with which they approach that peak and descend from it. If it were the case that the shape of individual trajectories displayed little or no variability across individuals, as argued by Hirschi and Gottfredson [51, 52] and by Gottfredson and Hirschi [39], this would not be a meaningful advantage.²² In the California Youth Authority data set analyzed here, however, there is substantial cross-individual variability. For example, among our black subjects released in 1986, the ratio of the estimated coefficient of age to the estimated coefficient of age-squared has a mean of -35.34 (When multiplied by $-1/2$, this ratio equals the age of peak criminality). Computing this ratio for all the subjects in this cohort, the lowest age of peak criminal involvement is seen to be 15.58; the highest is 48.91. Given that the highest age of any subject in the black 1986 cohort is 38, this subject has an arrest history that continues to rise up to the end of the data collection period. As it happens, only 1 subject of the 532 in the black 1986 cohort has an arrest history that continues to increase up to age 38, indicating how rare such a pattern is.

That everyone else in this cohort has a trajectory that peaks and then declines suggests that in this cohort, at least, there are virtually no life-course persistent offenders.

It is desirable to be able to take this variability in the shape of the age distribution of crimes into account in a flexible manner. The finite mixture model imposes excessively rigid constraints on the variability of the intercept and slopes by forcing the researcher to predict the group rather than the individual parameters that characterize the different trajectories.²³

It has been argued [85, 97] that a group-based approach may be better-suited to the study of phenomena (such as the onset of depression) that do not tend to follow a smooth path of temporal development. While it is true that psychiatric symptoms may occur, disappear, and reappear abruptly and at unpredictable moments rather than in the smooth manner characteristic of vocabulary acquisition in small children (for example), both approaches will, in fact, be equally ill-suited to modeling them because both rely on a parametric representation of the trajectories. In these circumstances, a multilevel pooled time-series cross-section analysis or an event history analysis may be more appropriate.

²² The group-based analysis of city homicide rates conducted by McCall et al. [66] illustrates this pattern with aggregate data. Their four trajectories display different levels of crime, but the temporal dependence of the crime rates in their data set is consistent with a single function of time and a random component to the level, i.e., to the intercept.

²³ Ezell and Cohen [30, p. 226] transcend this limitation by introducing interaction terms between classes and the individual age coefficients. This provides greater flexibility in modeling though at the cost of greater model complexity. We have not seen this method used in subsequent work.

Phillips and Greenberg [93] and Greenberg and Phillips [45] have argued that this may be the case when studying aggregate crime rates, which may rise and fall erratically from year to year. In this circumstance, a growth curve analysis may be a poor way to model temporal change.²⁴ McCall et al. [66] required a quintic polynomial to reproduce homicide rates in US cities, in a finite mixture modeling effort whose main conclusions are that city homicide rates in the years studied were moving largely in parallel, and that some cities have higher rates of homicide than others. These conclusions could also have been reached with a panel model including fixed effects for years, and interaction terms between city dummies and linear and quadratic terms for year.

Little of the statistical literature on finite mixture models discusses variance explained.²⁵ A simple way of assessing this would be to compute the correlation between the observed outcome at each age or time point, and the value of the outcome variable at that age predicted by the most probable trajectory. This value of R^2 could be compared with the reduction in variance accomplished by the introduction of level-2 variation in a multilevel model. This comparison might be an additional tool for comparing the approaches to trajectory analysis.²⁶

Thus far, I have been assuming that the central purpose of the researchers is to identify the true character of individual trajectories, and to determine the factors that influence these trajectories. However, this is not always the case. There are two other purposes for which the analyses of criminal careers might be undertaken. One is administrative. Agencies sometimes seek to classify a set of individuals into a finite number of mutually exclusive discrete categories for the purposes of doing something distinctive to those in each group, such as releasing an individual before trial or on parole, or making assignments to treatment groups.

The use of finite mixture modeling for administrative purposes does not pose the kinds of statistical problems we have delineated in this paper. In principle, it could be useful in such applications. There may, however, be other objections. Nagin and Tremblay [82] have argued that at present, our ability to classify individuals into groups and to use group membership to predict individuals' future criminality is too limited to warrant its use for this purpose. This limitation is equally present in predictions based on multilevel modeling. Objections to the use of inaccurate predictions in criminal justice decision-making have been voiced for some decades [1, 46, 119], and require no repetition here. Notwithstanding these objections, predictive methods are spreading in a range of criminal justice and civil proceedings, such as pretrial release and civil commitment proceedings for "sexually violent predators." In applications where these objections are not relevant, finite mixture models could be used whether or not there are discrete groups. At the same time, they do not offer striking advantages. In the multilevel modeling approach, one could still set cut-off points that would permit the division of a set of continuous trajectories into distinct groups.

²⁴ The quintic polynomial provides no insights into the reasons why homicide rates were moving upward or downward in the time period studied. The equation may simply be modeling upward and downward movements that are largely random.

²⁵ This omission in criminological literature is not unusual in criminological literature on crime causation [123].

²⁶ There are, however, conceptual issues that must be addressed when choosing a measure of R^2 to assess variance explained by a multilevel model [108].

The finite mixture modeling approach has also been touted as a way of representing a continuous distribution for purposes of pictorial representation. A graphical representation of a small number of group trajectories enables viewers to capture the main features of a data set at a glance by depicting typical trajectories. This may well be so. At the same time, nothing bars the researcher from plotting the trajectories of arbitrarily or randomly selected individuals to show the variability of actual trajectories. Smoothing techniques could be used to make these trajectories easier to grasp visually.

Whether smoothing is in fact a good idea must be considered in light of the purpose of the representation. Actual sequences of crimes or arrests tend to be somewhat erratic, reflecting the episodic character of a good deal of criminal behavior. Smoothing may present a picture that is pretty and easy to grasp, but at the possible price of obscuring an important feature of the patterns. Indeed, the “typical” trajectory identified by the group-based method may be anything but typical [21].

Moreover, even when researchers are careful to emphasize that the groups identified by a finite mixture modeling approach should not be fetishized, but only represent a simplified characterization of a more complex reality, there is a danger that listeners, viewers, and readers will pay insufficient attention to these cautionary remarks, and treat the groups as genuine [29, 42, 59, 110]. To their credit, proponents of the group-based approach have cautioned against the interpretation of the groups identified by the approach as “real.” It has been our experience that many listeners tune out these cautionary remarks. They attend conference presentations at which speakers voice the appropriate qualifications, yet they come away talking about “adolescent-limited offenders” and “life-course persistent offenders” as two genuinely distinct types of offenders, not as a convenient short-hand for different portions of a continuum. Indeed, authors of scholarly research articles lapse into this usage with a frequency that is high enough to be worrisome. They write as if the number of trajectories in their best-fitting models is a feature of the underlying data-generating process, not merely an approximation to a continuous distribution. For example, Weisburd et al. [122] speak of the number of trajectories in their model of longitudinal crime patterns in Seattle as a feature of the crime patterns that their analysis has revealed. Similarly, McCall et al. [66] “find” that there are four distinct trajectories of US city homicide rates without showing readers evidence that the observed trajectories actually cluster into distinct groups. Consequently, while we do not want to dismiss the value of finite mixture modeling as a tool for summarizing patterns of criminal careers, we are concerned that it may convey more than its users intend. For this reason, we think it is especially important that researchers not adopt the group-based approach as their first-choice method, but instead keep it in reserve for those occasions when a multilevel or ITS analysis shows trajectories that cluster into relatively discrete groups.

Acknowledgments I am grateful to Michael Ezell for sharing a portion of his data, and for advice and suggestions regarding the statistical analysis and the presentation of the research. I am also grateful to the anonymous referees for their comments on earlier drafts. An earlier version of this paper was presented to the annual meetings of the American Sociological Association and the American Society of Criminology in 2009.

References

1. American Friends Service Committee Working Party. (1971). *Struggle for justice: a report on crime and punishment in America*. New York: Hill and Wang.
2. Andruff, H., Carraro, N., Thompson, A., & Gaudreau, P. (2009). Latent class growth modelling: a tutorial. *Tutorial in Quantitative Methods for Psychology*, 5, 11–24.
3. Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32, 159–188.
4. Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338–363.
5. Bauer, D. J., & Curran, P. J. (2003). Overextraction of latent trajectory classes: much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Bulletin*, 8, 384–393.
6. Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological Methods*, 9, 3–29.
7. Becker, H. (1953). Becoming a marijuana user. *American Journal of Sociology*, 59, 235–242.
8. Becker, P., & Wetzell, R. F. (Eds.). (2006). *Criminals and their scientists: the history of criminology in international perspective*. New York: Cambridge University Press.
9. Beirne, P. (1993). *Inventing criminology: essays on the rise of homo criminalis*. Albany: State University Press of New York.
10. Beirne, P. (1994). *The origins and growth of criminology: essays on intellectual history, 1760–1945*. Brookfield: Dartmouth.
11. Benaglia, T., Hunter, D. R., Chauveau, D., & Young, D. S. (2009). Mixtools: an R package for analyzing finite mixture models. *Journal of Statistical Software*, 32, 1–29.
12. Berk, R. (2010). What you can and can't properly do with regression. *Journal of Quantitative Criminology*, 26, 481–487.
13. Blokland, A. A. J., & Neuwbeerta, P. (2005). The effects of life circumstances on longitudinal trajectories of offending. *Criminology*, 43, 1203–1240.
14. Blumstein, A., Cohen, J., Roth, J., & Visser, C. A. (Eds.). (1986). *Criminal careers and "career criminals"*. Washington, DC: National Academy Press.
15. Blumstein, A., & Rosenfeld, R. (1998). Explaining recent trends in U.S. homicide rates. *Journal of Criminal Law and Criminology*, 8, 1175–1216.
16. Blumstein, A., & Moitra, S. (1980). The identification of "career criminals" from "chronic offenders" in a cohort. *Law Policy Quarterly*, 2, 321–334.
17. Brame, R., Bushway, S., & Paternoster, R. (1999). On the use of panel research designs and random effects models to investigate static and dynamic theories of criminal offending. *Criminology*, 37, 599–641.
18. Brame, R., Nagin, D. S., & Wasserman, L. (2006). Exploring some analytical characteristics of finite mixture models. *Journal of Quantitative Criminology*, 22, 31–59.
19. Brame, R., Paternoster, R., & Piquero, A. R. (2012). Thoughts on the analysis of group-based developmental trajectories in criminology. *Justice Quarterly*, 29, 469–490.
20. Bushway, S., Brame, R., & Paternoster, R. (1999). Assessing stability and change in criminal offending: a comparison of random effects, semi-parametric, and fixed effect modeling strategies. *Journal of Quantitative Criminology*, 15, 23–61.
21. Bushway, S., Sweeten, G., & Nieuwbeerta, P. (2009). Measuring long term individual trajectories of offending using multiple methods. *Journal of Quantitative Criminology*, 25, 259–286.
22. Butler, S. M., & Louis, T. A. (1992). Random effects models with non-parametric priors. *Statistics in Medicine*, 11, 1981–2000.
23. Chang, I. (2005). *Bayesian Inference on Mixture Models and their Application*. Ph.D. Dissertation, Texas A & M University. Available at <http://repository.tamu.edu/bitstream/handle/1969.1/3990/etc.-tamu-2005A-STAT-Chan.pdf?sequence=1>.
24. Cheng, M.-Y., & Hall, P. (1998). Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society, Series B*, 60, 579–589.
25. Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. R. Yarrow (Eds.), *Handbook of statistical modeling in the social and behavioral sciences* (pp. 311–359). New York: Plenum.

26. Cohen, P., Chen, H., Hamigami, F., Gordon, K., & McCordle, J. J. (2000). Multilevel analyses for predicting sequence effects of financial and employment problems on the probability of arrest. *Journal of Quantitative Criminology*, 16, 223–235.
27. D'Unger, A. V., Land, K. C., McCall, P. L., & Nagin, D. S. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses of the London, Philadelphia, and Racine cohort studies. *American Journal of Sociology*, 103, 1593–1630.
28. Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminological trajectories. *Journal of Quantitative Criminology*, 20, 1–26.
29. Erosheva, E. A., Matsueda, R. L., & Telesca, D. (2014). Breaking bad: two decades of life-course data analysis in criminology, developmental psychology, and beyond. *Annual Review of Statistics and its Application*, 1, 301–332.
30. Ezell, M. E., & Cohen, L. E. (2005). *Desisting from crime: continuity and change in long-term crime patterns of serious chronic offenders*. New York: Oxford University Press.
31. Farrington, D. P. (1986). Age and crime. In M. Tonry & N. Morris (Eds.), *Crime and justice: an annual review of research* (pp. 189–250). Chicago: University of Chicago Press.
32. Farrington, D. P. (1995). The development of offending and antisocial behaviour from childhood: Key findings from the Cambridge Study in delinquent development. *Journal of Child Psychology and Psychiatry*, 36, 929–964.
33. Feldman, B. J., Masyn, K. E., & Conger, R. D. (2009). New approaches to studying problem behavior: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology*, 45, 652–676.
34. Francis, B., Soothill, K., & Fligelstein, R. (2004). Patterns of offending behavior: a new approach to typologies of crime. *European Journal of Criminology*, 1, 47–80.
35. Gibson, M. (2002). *Born to crime: Cesare Lombroso and the origins of biological criminology*. Westport: Praeger.
36. Glaeser, E. L., & Sunstein, C. R. (2009). Extremism and social learning. *Journal of Legal Analysis*, 1, 263–324.
37. Glueck, S., & Glueck, E. T. (1930). *500 criminal careers*. New York: Knopf.
38. Glueck, S., & Glueck, E. T. (1937). *Later criminal careers*. New York: Commonwealth Fund.
39. Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford: Stanford University Press.
40. Greenberg, D. F. (1977). Delinquency and the age structure of society. *Contemporary Crisis*, 1, 189–224.
41. Greenberg, D. F. (1985). Age, crime, and social explanation. *American Journal of Sociology*, 91, 859–866.
42. Greenberg, D. F. (1991). Modeling criminal careers. *Criminology*, 29, 17–46.
43. Greenberg, D. F. (2008). Age, sex, and racial distributions of crime. In E. Goode (Ed.), *Out of control: assessing the general theory of crime* (pp. 38–48). Stanford: Stanford University Press.
44. Greenberg, D. F., & Larkin, N. (1985). Age-cohort analysis of arrest rates. *Journal of Quantitative Criminology*, 1, 227–241.
45. Greenberg, D. F., & Phillips, J. A. (2012). Hierarchical linear modeling of growth curve trajectories using HLM. In G. David Garson (Ed.), *Hierarchical linear modeling: guide and applications* (pp. 219–247). Thousand Oaks: Sage.
46. Harcourt, B. (2006). *Against prediction: profiling, policing, and punishing in an actuarial age*. Chicago: University of Chicago Press.
47. Hartigan, J. A. (1985). Algorithm AS 217: computation of the dip statistic to test for unimodality. *Applied Statistics*, 34, 320–325.
48. Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *Annals of Statistics*, 13, 70–84.
49. Henderson, D. J. (2010). A test for multimodality of regression derivatives with application to nonparametric growth regressions. *Journal of Applied Econometrics*, 25, 458–2010.
50. Hilbe, J. (2007). *Negative binomial regression*. New York: Cambridge University Press.
51. Hirschi, T., & Gottfredson, M. R. (1983). Age and the explanation of crime. *American Journal of Sociology*, 89, 552–584.
52. Hirschi, T., & Gottfredson, M. R. (1985). Age and crime, logic and scholarship: comment on Greenberg. *American Journal of Sociology*, 91, 22–27.
53. Horney, J., Osgood, D. W., & Marshall, I. H. (1995). Criminal careers in the short-term: intra-individual variability in crime and its relation to local life circumstances. *American Sociological Review*, 60, 655–673.
54. Jennings, W. G., & Reingle, J. M. (2012). On the number and shape of developmental/life-course violence, aggression, and delinquency trajectories: a state-of-the-art review. *Journal of Criminal Justice*, 40, 472–489.

55. Jones, B., Nagin, D. S., & Roeder, K. (2001). A SAS Procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research*, 29, 374–393.
56. Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.
57. Kass, R. E., & Wasserman, R. E. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90, 928–934.
58. Kreuter, F., & Muthén, B. (2008). Analyzing criminal trajectory profiles: bridging 'multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology*, 24, 1–31.
59. Laub, J. H., & Sampson, R. J. (2003). *Shared beginnings, divergent lives*. Cambridge: Harvard University Press.
60. Laub, J. H., Nagin, D. S., & Sampson, R. J. (1998). Trajectories of change in criminal offending: good marriages and the desistance process. *American Sociological Review*, 63, 225–238.
61. Lee, K.-J., Chen, R.-B., & Wu, Y. N. (2016). Bayesian variable selection for finite mixture model of linear regressions. *Computational Statistics and Data Analysis*, 95, 1–16.
62. Leisch, F. (2004). FlexMix: a general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11, 1–18.
63. Liu, M. (2011). *Using latent profile models and unstructured growth mixture models to assess the number of latent classes in growth mixture modeling*. Ph.D. Dissertation, University of Maryland.
64. Lynch, K. G., Nagin, D. S., & Roeder, K. (1999). Modeling uncertainty in latent class membership: a case study in criminology. *Journal of the American Statistical Association*, 94, 766–776.
65. Macmillan, R. (2008). Review. *Contemporary Sociology*, 37, 159–160.
66. McCall, P. L., Land, K. C., & Parker, K. F. (2011). Heterogeneity in the rise and decline of city-level homicide rates, 1976–2005: a latent trajectory analysis. *Social Science Research*, 40, 363–378.
67. McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. Hoboken: Wiley.
68. McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
69. McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
70. Meehl, P. E. (1999). Clarifications about taxometric method. *Applied and Preventive Psychology*, 8, 165–174.
71. Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463–469.
72. Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus User's Guide*. 5th and Muthén.
73. Muthén, L. K., & Asparouhov, T. (2008). Growth mixture modeling: analysis with non-gaussian random effects. In M. G. Verbeke & G. Molenberg (Eds.), *Longitudinal data analysis* (pp. 147–165). Boca Raton: Chapman & Hall/CRC Press.
74. Nagin, D. S. (1999). Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychological Methods*, 4, 139–157.
75. Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge: Harvard University Press.
76. Nagin, D. S. (2007). Overview of a semi-parametric, group-based approach for analyzing trajectories of development. In D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman (Eds.), *Cambridge handbook of violent behavior and aggression* (pp. 740–749). New York: Cambridge University Press.
77. Nagin, D. S. (2010). Group-based trajectory modeling: an overview. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 53–68). New York: Springer.
78. 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.
79. Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling (nearly) two decades later. *Journal of Quantitative Criminology*, 26, 445–453.
80. Nagin, D. S., & Paternoster, R. (1991). On the relationship of past to future participation in crime. *Criminology*, 29, 163–189.
81. Nagin, D. S., & Paternoster, R. (1992). The onset and persistence of offending. *Criminology*, 30, 501–523.
82. Nagin, D. S., & Tremblay, R. E. (2005). Developmental trajectory groups: fact or a useful statistical fiction? *Criminology*, 43, 873–904.
83. Nagin, D. S., & Tremblay, R. E. (2005). From seduction to passion: a response to Sampson and Laub. *Criminology*, 43, 915–918.
84. Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *The Annals of the American Academy of Political and Social Science*, 602, 145–154.

85. Nagin, D. S., & Tremblay, R. E. (2005). Further reflections on modeling and analyzing developmental trajectories: a response to Maughan and Raudenbush. *The Annals of the American Academy of Political and Social Science*, 602, 145–154.
86. Neuhaus, J. M., & McCulloch, C. E. (2008). *The consequences of misspecifying the random effects distribution when fitting generalized linear mixed models*. Paper presented to the 2008 meeting the Stata User's Group, Washington, DC. Available at www.stata.com/meeting/fnasug08/neuhaus_Stata2008.talk.pdf. Last accessed 14 July 2015.
87. Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of latent classes in class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535–569.
88. Patterson, G. R., Debaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial-behavior. *American Psychologist* 44, 329–335.
89. Patterson, G. R., Forgatch, M. S., Yoerger, K. L., & Stoolmiller, M. (1998). Variables that initiate and maintain an early-onset trajectory for juvenile offending. *Development and Psychopathology* 10, 531–547.
90. Paternoster, R., Bushway, S., Brame, R., & Apel, R. (2003). The effect of teenage employment on delinquency and problem behaviors. *Social Forces*, 82, 297–335.
91. Petras, H., & Masyn, P. (2010). General growth mixture analysis with antecedents and consequences of change. In H. Petras & D. Weisburd (Eds.), *Handbook of quantitative criminology*. New York: Springer.
92. Phenix, D. (2010). *Criminal careers: the distribution of change*. Ph.D. Dissertation, New York University.
93. Phillips, J., & Greenberg, D. F. (2008). A comparison of methods for analyzing criminological panel data. *Journal of Quantitative Criminology*, 24, 51–72.
94. Piquero, A. R. (2008). Taking stock of developmental trajectories of criminal activity over the life course. In A. M. Liberman (Ed.), *The long view of crime: a synthesis of longitudinal research* (pp. 23–78). New York: Springer.
95. Piquero, A. R., Farrington, D. P., & Blumstein, A. (2007). *Key issues in criminal career research: new analyses of the Cambridge study in delinquent development*. New York: Cambridge University Press.
96. Rafter, N. H. (1997). *Creating born criminals*. Urbana: University of Illinois Press.
97. Raudenbush, S. W. (2005). How do we study what happens next? *The Annals of the American Academy of Political and Social Science*, 602, 131–144.
98. Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.
99. Roeder, K., Lynch, K. G., & Nagin, D. S. (1999). Modeling uncertainty in latent class membership: a case study in criminology. *Journal of the American Statistical Association*, 94, 766–776.
100. Rowe, D. C., Wayne Osgood, D., & Alan Nicewander, W. (1990). A latent trait approach to unifying criminal careers. *Criminology*, 28, 237–270.
101. Sampson, R. J., & Laub, J. H. (2003). Life-course desisters? Trajectories of crime among the delinquent boys followed to age 70. *Criminology*, 41, 301–339.
102. Sampson, R. J., & Laub, J. H. (2005). Seductions of method: rejoinder to Nagin and Tremblay's 'developmental trajectory groups: fact or fiction?'. *Criminology*, 43, 905–913.
103. Schork, N. J., & Schork, M. A. (1988). Skewness and mixtures of normal distributions.
104. Shaw, C. R. (1930). *The jack-roll; a delinquent boy's own story*. Chicago: University of Chicago Press.
105. Shaw, C. R. (1931). *The natural history of a delinquent career*. Chicago: University of Chicago Press.
106. Skardhamar, T. (2009). Reconsidering the theory on adolescent-limited and life-course persistent anti-social behavior. *British Journal of Criminology*, 49, 863–878.
107. Skardhamar, T. (2010). Distinguishing facts and artifacts in group-based modeling. *Criminology*, 48, 295–320.
108. Steele, R. (2015). Model selection for multilevel models. In M. A. Scott, J. S. Simonoff, & B. D. Marx (Eds.), *The SAGE handbook of multilevel modeling: model selection for multilevel models* (pp. 109–127). Thousand Oaks: Sage.
109. Steffensmeier, D. S., Allen, E. A., Harer, M. D., & Streifel, C. (1989). Age and the distribution of crime. *American Journal of Sociology*, 94, 803–831.
110. Sterba, S. K., Baldasaro, R. E., & Bauer, D. J. (2012). Factors affecting the adequacy and preferability of semiparametric group-based approximations of continuous growth trajectories. *Multivariate Behavioral Research*, 47, 590–634.
111. Sunstein, C. (2002). The law of group polarization. *Journal of Political Philosophy*, 10, 175–195.
112. Tewksbury, R., & Jennings, W. G. (2010). Assessing the impact of sex offender registration and community notification on sex-offending trajectories. *Criminal Justice and Behavior*, 37, 570–582.

113. Titterton, D. M., Smith, A. F. M., & Makow, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
114. Tong, X., & Zhanag, Z. (2012). Diagnostics of robust growth modeling using Student's t distribution. *Multivariate Behavioral Research*, 47, 493–518.
115. Uggen, C., & Thompson, M. (2003). The socioeconomic determination of ill-gotten gains: within-person changes in drug use and illegal earnings. *American Journal of Sociology*, 109, 146–185.
116. Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91, 217–221.
117. Verbeke, G., & Lessafre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*, 232, 541–556.
118. Vermunt, J. K., & Magidson, J. (2000). *Latent gold user's guide*. Belmont: Statistical Innovations, Inc.
119. von Hirsch, A. (1976). *Doing justice: the choice of punishments*. New York: Hill and Wang.
120. Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, 10, 635–656.
121. Warren, J. R., Luo, L., Halpern-Manners, A., Raymo, J. M., & Palloni, A. (2015). Do different methods for modeling age-graded trajectories yield consistent and valid results? *American Journal of Sociology*, 20, 1809–1856.
122. Weisburd, D., Bushway, S., Lum, C., & Yang, S.-M. (2004). Trajectories of crime at places: a longitudinal study of street segments in the City of Seattle. *Criminology*, 42, 283–322.
123. Weisburd, D., Cave, B., & Piquero, A. (2015). How do criminologists interpret statistical explanation of crime?: a review of quantitative modeling in published studies. In A. Piquero & C. Wellford (Eds.), *Handbook of Criminological Theory*. New York: Springer Verlaag.
124. Wolfgang, M. E., Figlio, R. M., & Sellin, T. (1971). *Delinquency in a birth cohort*. Chicago: University of Chicago Press.