



## How Many Latent Classes of Delinquent/Criminal Careers? Results from Mixed Poisson Regression Analyses

Amy V. D'Unger; Kenneth C. Land; Patricia L. McCall; Daniel S. Nagin

*The American Journal of Sociology*, Vol. 103, No. 6. (May, 1998), pp. 1593-1630.

Stable URL:

<http://links.jstor.org/sici?sici=0002-9602%28199805%29103%3A6%3C1593%3AHMLCOD%3E2.0.CO%3B2-7>

*The American Journal of Sociology* is currently published by The University of Chicago Press.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# How Many Latent Classes of Delinquent/ Criminal Careers? Results from Mixed Poisson Regression Analyses<sup>1</sup>

Amy V. D'Unger and Kenneth C. Land  
*Duke University*

Patricia L. McCall  
*North Carolina State University*

Daniel S. Nagin  
*Carnegie Mellon University*

This article reviews questions about different categories of criminal careers, summarizes Poisson latent class regression models, describes procedures for evaluating the optimal number of latent classes, and applies this methodology to data from male cohorts taken from the cities of London, Philadelphia, and Racine. Four latent classes of offending careers is an appropriate number for the London cohort, but five classes can be justified for the Philadelphia data. In the case of the Racine cohorts, five classes may be detected for the 1942 and 1955 cohorts but only four for the 1949 cohort. Despite the varying numbers of latent offending classes, there clearly is a small number of typical age patterns.

The Poisson distribution and Poisson regression models are fast becoming key methodological tools for the development of micromodels of delinquent/criminal careers in the criminological literature (Blumstein et al. 1986; Greenberg 1991; Nagin and Land 1993; Land, McCall, and Nagin 1996; Land and Nagin 1996). Mixed Poisson regression models, in which the mixing distribution is nonparametric, and associated semiparametric maximum-likelihood estimators, have yielded interesting results in the modeling of this type of career. One property of such models is that small numbers of *latent classes* (categories, groupings) of individuals can

<sup>1</sup> The research reported here was supported, in part, by the National Science Foundation (grants SBR-9511490 and SBR-9511412). The authors thank Lawrence Cohen, Michael Ezell, and the *AJS* reviewers for helpful comments in earlier drafts. Direct correspondence to Amy V. D'Unger, Department of Sociology, Duke University, Durham, North Carolina 27708-0088.

be identified with respect to similarity in patterns or careers of delinquent and criminal offending over age, net of regression effects of other socio-demographic characteristics that contribute to delinquent/criminal propensity.<sup>2</sup> Researchers studying delinquent/criminal offending over time (e.g., Rowe, Osgood, and Nicewander 1990; Sampson and Laub 1993, 1995) have applied key concepts and methods from life-course researchers such as Elder (1975, 1992), Caspi (Caspi and Bem 1990; Caspi and Moffitt 1993), and Abbott and Hrycak (1990) to form a new perspective on how offending should be studied. In particular, the concepts of continuity versus change and of trajectories and transitions have been salient for studying the pathways into crime. These life-course concepts and methods have helped researchers in thinking about how both juveniles and adults enter into a pattern of delinquent/criminal offending, as well as how they desist from such offending. Mixed Poisson regression models, with their ability to identify latent groups of offenders with similar offending trajectories, are a relatively new tool for studying continuity and change in offending patterns at the individual level and can advance criminological knowledge of offending over the life course.

But the utility of Poisson latent class models goes well beyond the field of criminology. Indeed, researchers in other areas both within and beyond sociology (e.g., psychology, demography, economics) can benefit from the insights of this methodological technique; in particular, those who are concerned about the *hidden or unobserved heterogeneity* that may exist in either cross-sectional or longitudinal event count data. Those working within a life-course framework, in which trajectories of repeated events are studied over time across a sample of individuals, will find this technique especially helpful. For example, these models could be fruitfully applied to the study of the publication careers of academics (where the dependent variable is the number of publications per year or other time period) over the course of their life span. Similarly, the frequency of a particular sexual act, such as unprotected sexual intercourse, could be studied for a cohort of individuals over time. The application of these models might lead to the discovery of latent classes where only a single observed class had been previously assumed and, therefore, theorized and empirically studied. While the focus of the present article is on the application of semiparametric mixed Poisson regression analysis to the study of

<sup>2</sup> Conventional latent structure/latent class analysis is a technique for analyzing the latent structure of qualitative/categorical variables (see, e.g., Goodman 1974a, 1974b, 1987; Clogg 1994). This technique (analogous, in various ways, to factor analysis for quantitative variables) can be used to determine whether the observed relationships among several categorical manifest variables can be explained in terms of a smaller number of latent variables. For a recent application of latent structure analysis to job characteristics of men and women, see Birkelund, Goodman, and Rose (1996).

delinquent/criminal careers, readers interested in other substantive problems within sociology may find the application of this new analytic tool of interest as well.

In analysis (Nagin and Land 1993) of the London cohort data of West and Farrington (1973, 1977) and in recent analysis (Land et al. 1996) of the 1958 Philadelphia birth cohort data of Tracy, Wolfgang, and Figlio (1990), four distinct classes of careers have emerged, ranging from non-offenders to high-rate offenders with sustained criminal careers. However, the question of whether four categories of careers are statistically optimal has not been systematically addressed in these previous works nor have statistical methods of comparing various models been explicated and applied. Similarly, previous studies have not addressed the question of whether the inference of the number of latent classes depends on the sample size of the cohort study or the question of how generalizable the results are across samples and types of communities. Finally, the implications of these multiple categories for law enforcement, crime control, and public policy have not been fully developed. For example, how might intervention programs be differently structured, given differing developmental trajectories of offending (e.g., chronic offenders vs. offenders whose activity peaked in adolescence)? The implications of this analysis have not yet been linked to sociological or criminological theory or public policy. While these questions are beyond the scope of an individual article, they can only be answered when the optimal numbers of latent categories of offenders are determined. For these reasons, it is unclear whether researchers should proceed with the assumption that the four-category model of delinquent/criminal careers found in previous analyses is both methodologically and theoretically sound.

This question of how many latent classes of criminal careers are optimal, and why the number of categories itself is important, has gained salience for criminological theory in light of recent theoretical debates between proponents of self-control theory (Hirschi and Gottfredson 1995) and those who emphasize the importance of social bonds that develop over the life course as factors in crime causation (Sampson and Laub 1995). Much of the previous research on criminal careers used one class of offenders, thus only making distinctions between offenders and non-offenders. In doing so, the heterogeneity in offending trajectories or patterns is lost and general theories such as that of Hirschi and Gottfredson (1995) are presumed to apply to all offenders. Past research on crime has also been conducted largely with the use of cross-sectional data. Newer models will allow those who are advocates of longitudinal data collection and analysis (e.g., Sampson and Laub 1993) to better map the developing trajectories of any type of behavior, including offending, thus leading to theoretical refinements.

In particular, semiparametric mixed Poisson regression models can help to resolve this theoretical debate. Indeed, a recent review of this debate by Cohen and Vila (1996) suggests that distinctions among qualitatively different categories along the criminal continuum may be a key to further theoretical synthesis and advances. In an attempt to clarify these and related issues, this article reviews some theoretical questions recently raised about the number and characteristics of different categories of criminal careers, summarizes the basic features of mixed Poisson regression models for delinquent and criminal careers, describes statistical testing procedures for evaluating the optimal number of latent classes, and reports results of applications of this methodology to data from male cohorts taken from the cities of London, Philadelphia, and Racine. In addition to providing an illustration of how semiparametric models may be applied within the field of criminology, general life-course concepts will also be introduced and integrated with both the methodological technique and the substantive example, demonstrating how these models may be applied to other topics of interest to sociologists.

#### QUALITATIVELY DISTINCT CRIMINAL CAREER CATEGORIES AND LIFE-COURSE TRAJECTORIES

The concept of the criminal career, broadly conceived as the longitudinal sequence of delinquent and criminal acts committed by an individual, is noncontroversial within the field of criminology. Controversies arise, however, over the question of whether criminals are a distinctive category in the population (as in classical criminal careers models; see Blumstein et al. 1986), whether criminal propensities are relatively constant across the life span (as in the self-control theory of Gottfredson and Hirschi [1990]), and whether qualitative changes in social bonds associated with life-course transitions can change the continuity or persistence in offending (as in the life-course perspective of Sampson and Laub [1993]). Because of the nature of these questions, the life-course perspective is one with much utility for the study of criminal careers.

Life-course sociology, according to Elder (1992, p. 1121) is both "a concept and a theoretical perspective." It refers to both the "age-graded life patterns embedded in social institutions and subject to historical change" and a "field of inquiry . . . providing a framework that guides research" (Elder 1992, p. 1121). Life-course sociologists study trajectories, or patterns of behavior (e.g., in work, family, schooling, marital status, offending, etc.) that extend through the individual's life, as well as how major life changes or transitions affect these behavioral trajectories. These transitions usually involve a change in the individual's role in society, such as the transition to adulthood, parenthood, or a major career shift.

The life course is given meaning through the age structure of society, as well as age norms, sanctions, social timetables for certain behaviors, and historical context (Elder 1975).

Studying any subject from a life-course perspective thus implies the use of longitudinal data that follows individuals over time. In criminology, this has been difficult due to a lack of available resources, which has hampered the development of testable theories of crime and delinquency. As Sampson and Laub (1993, p. 23) have pointed out, "Criminology has been dominated by narrow sociological and psychological perspectives, coupled with a strong tradition of research using cross-sectional data on adolescents." This combination of limited theoretical perspective and methodological technique has particularly hampered our ability to understand the criminal career, which is both longitudinal and dynamic in nature. In addition, it has led to a debate between those who advocate the *life-course perspective* (Sampson and Laub 1993) and those who advocate the notion of *stability in criminal offending over time* (Gottfredson and Hirschi 1990).

In a recent review of the self-control versus life-course variability debate (Hirschi and Gottfredson 1995; Sampson and Laub 1995), Cohen and Vila (1996) articulate the *paradox of persistence* in which consistent empirical evidence shows delinquent behavior in children to be one of the best predictors of criminal behavior in adults, while at the same time other empirical evidence consistently shows that most delinquents do not go on to become adult offenders. According to Cohen and Vila, this paradox can be explained by taking a closer look at the data on crime, which shows that there is far less stability in criminal behavior over the life course than previous research (and theory) has acknowledged. Indeed, these authors argue that "criminality appears to be highly stable only among a relatively small number of males whose antisocial behavior is relatively extreme" (Cohen and Vila 1996, p. 23). In addition, they note that individuals at this most extreme end of the criminality continuum appear to share many attributes commonly described in the psychological literature as indicators of *sociopathy* (also known as antisocial personality disorder), that empirical estimates suggest that this small group of sociopaths composes between 3% and 4% of the male population (and less than 1% of the female population), that they account for between 33% and 80% of the chronic offender population, and that they are responsible for about half of all the serious (felony) crimes that occur annually.

Cohen and Vila (1996) also suggest that, in order to move beyond the self-control versus life-course variability debate, it may be necessary to recognize the existence of at least a second category of offender career pattern composed of individuals on the high end of the criminal propensity continuum who are not as high in criminal propensity as those who exhibit extreme antisocial and sociopathic traits. This view is similar to

that of Moffitt (1993) who argues that a high propensity to engage in crime and other high-risk behaviors persists over the life course for some high-rate offenders (whom Moffitt calls *life-course persistents*) but that there also exists a second, larger category of individuals who are high-rate offenders only during adolescence (whom Moffitt calls *adolescence limiteds*). Those in this group, like the life-course persistent deviants, are heavily involved in crime and delinquency during their youth. Moffitt hypothesizes that the onset of delinquent behavior among the adolescence limiteds results from their mimicking the antisocial lifestyle of the life-course persistent youths. However, unlike offending by the life-course persistents, offending within the adolescence-limited subpopulation declines precipitously after about age 18. By comparison, Cohen and Vila (1996, p. 146) view these individuals who eventually desist as “competitively disadvantaged” offenders who are highly opportunistic but perhaps more tractable (e.g., susceptible to the influence of changes in social bonds associated with life-course transitions such as work or marriage) than the high-end offenders described in Gottfredson and Hirschi’s (1990) self-control theory. Cohen and Vila (1996, p. 146) also suggest that a “substantial” portion of all serious crimes are committed by this second type of offender.

The use of semiparametric mixed Poisson models can, to some degree, help to identify these classes of offenders over the span of multiple years, going beyond the relatively short span of adolescence typically used in much criminological research. In addition, this methodology has implications for the theoretical perspective from which crime must be studied. While the empirical analyses reported in this article cannot address all of the hypotheses advanced by Moffitt (1993) and Cohen and Vila (1996) about characteristics of the two postulated offender classes and the causal mechanisms that produce each type, we will begin to address some questions raised by these theoretical constructs and illustrate the use of latent class models in life-course research with a criminological example. First, using the recently developed semiparametric mixed Poisson regression model, we will address the question of how many latent classes (categories or groupings) of members of longitudinal cohort studies with respect to their offending careers are optimally identifiable in a well-defined statistical sense. Second, by analyzing three classical longitudinal studies of delinquent/criminal offending from three distinctive settings, we will begin to identify whether and how the measurement of the dependent variable (offending), the type of cohort studied, and the nature of the community from which it is drawn affects conclusions about the optimal number of categories. Third, we will sort—after estimation of the parameters of the optimal model for each cohort—longitudinal cohort members into the class of offenders to which they most probably belong. After this sorting, we then will plot and compare the expected (or predicted on the basis of

model parameters) age trajectories of offending of each of the categories of offenders. Finally, for each category of offenders, we will compute a number of summary statistics that can address hypotheses about the relative size and nature of the offender classes. From these analyses come implications for further study in the field of criminology, as well as other areas of sociology.

#### POISSON AND MIXED POISSON REGRESSION APPROACHES

Efforts to model longitudinal patterns of criminal offending, or criminal careers, have been hampered by the limitations of traditional models, such as the conventional Poisson and negative binomial regression models (Cameron and Trivedi 1986), for dealing with sequences of events that are rare occurrences. One limitation is that Poisson-based models, which include the negative binomial model, work under the assumption that events occur independently (conditional on the regressors) over time. Given the nature of criminal activity, this assumption of independence is most likely violated. The commission of a crime may raise the probability that an individual may commit a subsequent crime, called *state dependence* (Nagin and Paternoster 1991; Nagin and Farrington 1992b) or events/crimes may occur in spells that are not independent of each other. Such features of criminal careers make the traditional Poisson model and its assumption of independence inappropriate.

A second major limitation of the conventional *Poisson regression model* is its requirement of equality of the mean and variance of the offense rate conditional on the explanatory variables. This mean, denoted by the parameter  $\lambda_{it}$ , represents the expected rate at which specific (in the present case, delinquent or criminal) events are expected to occur for individual  $i$  per unit of time  $t$ , or the mean rate of occurrence. One way of relaxing the requirement of equality of the mean and variance in the conventional Poisson regression model is through the addition of an error term that allows for randomness in  $\lambda_{it}$ . Such a term can account for errors caused by misspecification of the regression model, such as through the omission of a relevant independent variable or other unexplained randomness found in  $\lambda_{it}$ . With the incorporation of such an error term, sources of both *observed* and *unobserved heterogeneity* may be taken into account (see Land et al. [1996] or Nagin and Farrington [1993b] for a discussion of unobserved or hidden heterogeneity).

The problem of unobserved heterogeneity is conventionally corrected through the specification of a gamma distribution (among members of the population from which a sample is drawn) for the error term in the regression model determining  $\lambda_{it}$ . The gamma family of distributions is a class of continuous distributions for random variables that take on only positive



or nonnegative values and includes the chi-square distribution as a special case. When the error term in the regression model for the  $\lambda_{it}$  is distributed in the population according to a gamma distribution, the random variable  $Y$ , or number of events per individual per unit of time, takes on a negative binomial distribution. The benefit of combining the Poisson distribution of events with a gamma distribution of errors for  $\lambda_{it}$  in order to yield a negative binomial distribution of events is the flexibility that one gains. Specifically, unlike in the Poisson regression model, the conditional variance of the negative binomial regression model is not constrained to be equal to its mean. While this combination still allows for the modeling of rare events under the assumptions of the Poisson distribution (events are rare within the total population and for most individuals), the addition of the gamma distribution of errors to create the *negative binomial regression model* allows individuals with identical values on the measured variables used as regressors to have expected rates of delinquent/criminal acts (i.e.,  $\lambda_{it}$ ) which are gamma distributed throughout the population rather than set to the same conditional mean rate.<sup>3</sup>

The development of *semiparametric mixed Poisson regression models* increases the flexibility of tools available to researchers for modeling patterns of criminal careers (Nagin and Land 1993; Land et al. 1996; Land and Nagin 1996). Semiparametric models generalize mixtures of Poisson distributions as discussed above. In Poisson mixture models, the error term is conceived of as unique to each individual in a longitudinal sample and invariant across time periods (i.e., over the life span of a delinquent/criminal). The negative binomial model assumes this "persistent unobserved (population) heterogeneity" (Nagin and Farrington, 1992a, 1992b) is gamma distributed whereas the semiparametric mixed Poisson models make no specific parametric assumptions about how the hidden heterogeneity is distributed throughout the population. Rather, it is approximated semiparametrically. Such a technique is desirable given that theory rarely will be strong enough to guide the specification of the distribution of the error term and that, even if theory were strong enough to do so, this choice may be incompatible with the data (Land et al. 1996). This innovation of the *semiparametric error term* combined with *parametric estimations of*

<sup>3</sup> When it is large, the Poisson distribution and the associated regression model can be more accurately approximated by the normal (Gaussian) distribution and the corresponding normal error regression model (cf. Land et al. 1996); therefore, Poisson regression is most frequently used in cases where the rate of event occurrence is rare. In the case of delinquent/criminal careers, Poisson regression is appropriate given that the mean numbers of events per individual per unit of time (e.g., contacts with the police, arrests, convictions, or similar measures of criminal offending per year of age) usually are small, creating a highly skewed distribution that is not well approximated by the normal distribution.

*the regression component* of the model using Poisson regression allows for more flexibility in the estimation of micromodels of criminal careers.<sup>4</sup>

Semiparametric mixed Poisson regression models presented in Nagin and Land (1993) utilize semiparametric maximum-likelihood (SPML) estimators to produce estimates of the parameters of the model. The likelihood function for this estimator was described in detail in Land et al. (1996); therefore, only the essentials of the approach will be sketched out here. Consider, for instance, a model that includes two points of support (which divide the sample into two latent classes or categories). The lambdas for the two groups of individuals may be estimated using the regression specification:

$$\ln \lambda_{it}^a = (\beta_0 + \epsilon_a) + \mathbf{X}_{it}\beta, \quad (1a)$$

$$\ln \lambda_{it}^b = (\beta_0 + \epsilon_b) + \mathbf{X}_{it}\beta, \quad (1b)$$

where the distribution of the mixing variable is approximated by two points of support, denoted by  $\epsilon_a$  and  $\epsilon_b$ ,  $\beta_0$  is the overall constant of the model (removed from the regression coefficient vector so that the adjustments to this constant can be explicitly identified), and  $\mathbf{X}_{it}$  and  $\beta$  are the matrix of regressor observations and regression coefficients from which the unit vector and intercept terms have been deleted, respectively. These two equations represent the rates of offending of individuals in the sample population whose hidden heterogeneity is approximated by one of the two points of support,  $a$  or  $b$ , corresponding to the specification of two latent groupings or categories of delinquent/criminal careers. The model itself generates appropriate numerical estimates for the points of support, as opposed to fixing points prior to estimation, by maximizing an appropriate likelihood for all of the data, including nonoffenders (individuals with no recorded offenses). The model also generates estimates of  $m_a$  and  $m_b$ , the proportions of the population at each point of support with the constraint that  $m_a + m_b = 1$ .

Utilizing the semiparametric mixed Poisson regression models and their associated SPML estimators, models with varying numbers of points of support may be estimated and compared against each other in an attempt to assess the optimal number of points of support and, therefore, categories of delinquent/criminal careers. This begs the question, however, of how to define "optimal" and appropriate statistical tests and decision rules for judging optimality. This question is addressed in the appendix by use of the Bayesian information criterion (BIC), which, under some conditions,

<sup>4</sup> Use of the term "semiparametric" to describe the models follows conventions described in Oakes (1988), who defines these models to be ones that include as unknowns both parameters and functions.

can be interpreted as a Bayes factor or index of the weight of the empirical evidence for one model as compared to another. As an overall criterion for making decisions about the optimal numbers of latent classes of offending careers in models for cohort studies, we will, following the guidelines described in the appendix, choose those models that minimize the BIC. In addition, however, we require that a chosen model have meaningful parameter estimates. This means, in particular, that the estimated parameters must be sufficiently uncorrelated that their standard errors can be computed.<sup>5</sup>

## DATA

Data for the present analyses are drawn from three cohort studies of the development of delinquent/criminal careers—the London panel study of David Farrington and Donald West (West and Farrington 1973, 1977), the 1958 Philadelphia cohort study by Marvin Wolfgang and associates (Tracy et al. 1990), and the Racine, Wisconsin, cohorts studied by Lyle W. Shannon (1988, 1991). Each of these studies has been sufficiently widely reported and studied in the criminological literature that it could be said to have attained the status of a classic cohort study. Each also has some unique features that permit the identification of commonalities as well as distinct features in answers to the question of the optimal number of categories of delinquent/criminal careers.

### The London Cohort

The London study is a prospective longitudinal survey of 411 mostly white males (born in 1953–54) from a stable and traditionally working-class area. Data collection began in 1961–62, when the boys were about eight years old. Criminal involvement is measured by *convictions for criminal offenses* for which data are available for all individuals in the sample through age 32, with the exception of eight individuals who died prior to this age. The conviction counts are for delinquent/criminal acts and do not include convictions for traffic offenses or for offenses deemed to be of minor seriousness (e.g., drunkenness or simple assault). The analy-

<sup>5</sup> For the computation of standard errors, our computer program inverts the Hessian matrix (the matrix of second-order partial derivatives of the log-likelihood function with respect to model parameters). If there is too much collinearity (linear dependence) among the columns (corresponding to estimated parameters) of a Hessian matrix, then it cannot be inverted. This implies that at least some of the estimated parameters are highly correlated and that one point of support is not sufficiently distinct from the others for the inverse to exist.

ses of this study reported here are based on data for the 403 individuals for whom there was a complete conviction history (recorded by year of age) between the tenth and the thirty-second birthdays. Of these, 36% had at least one conviction. For those who were convicted, the average number of convictions over the observation period was 4.4. One of the strengths of the London study is the impressive array of measurements on the cohort members—psychological characteristics, socialization variables, and family background. For a complete discussion of the data set, see West and Farrington (1973).

With regard to the community context of the London cohort, the total population of the city of London declined from about 3.5 million persons in 1961 to about 2.5 million in 1981. The urban section of London from which the cohort was sampled was characterized by overcrowded neighborhoods and a predominance of public housing, as well as a nearly all-white population.

### The Philadelphia Cohort

The 1958 birth cohort study in Philadelphia is a replication of an earlier 1945 birth cohort study of Wolfgang, Figlio, and Sellin (1972) and is composed of all subjects born in the year 1958 and residing in the city of Philadelphia from their tenth through eighteenth birthdays. A total of 13,160 males and 14,000 females were included in the cohort (57.2% non-white), data for which were drawn from birth and school records in the city of Philadelphia. Individuals in the sample were followed from ages 8 to 26 (1966–84). The main indicator of delinquency/crime up to age 18 was number of *contacts with the Juvenile Aid Division of the Philadelphia Police Department* by age. This contact did not have to come in the form of an arrest but may have been a less formal contact resulting in the release of the juvenile to his/her parent. After the age of 18, any subsequent police encounters recorded on regular police forms were used as data sources.

The comprehensiveness of the panel and the relatively unfiltered (by the justice system) measure of delinquency/criminal behavior are strengths of this study. On the other hand, the study does not include a large number of measures of the characteristics of the individual cohort members or their social environments for use as covariates in explanatory studies. In addition to recording the sex and race of each member, an individual-level measure of socioeconomic status is available. The socioeconomic status measure is derived from a principal components analysis of 10 indicators from the 1970 census of the social status of the juvenile's census tract. For a complete discussion of the data set, see Tracy et al. (1990).

Because the London cohort is all male, we limit the analyses of the

Philadelphia cohort reported below also to males, and, because the number of nonblack/nonwhite males in the cohort is quite small, only to those who are white or black. In addition, because the models applied herein are statistical models that efficiently use information from samples, it will not be necessary to include information from the entire cohort of males. On the other hand, the large size and comprehensiveness of this male cohort will allow us to perform some analyses of the sensitivity of inferences about the numbers of latent groupings of individuals with respect to age patterns of offending. Specifically, we draw random samples of sizes 500, 1,000, and 2,000 from the total number of (white and black) members of the male cohort for use in estimation of the models reported below. For the sample of size 1,000, we find that 39.3% had at least one official police contact for a felony or misdemeanor over ages 8–26. For those with one or more police contacts, the average number of contacts for these offenses over ages 8–26 was 4.7.

With regard to the community context of the Philadelphia cohort, it can be noted that—consistent with the experience of many large northeastern U.S. cities—the population of the city of Philadelphia declined in all decennial censuses since 1950. In particular, the population was recorded as 2 million in the 1960 census, 1.94 million in 1970, and 1.69 million in 1980. Of this total population, 26.7% were nonwhite in 1960, 34.4% in 1970, and 41.6% in 1980 (when it was separately estimated in the decennial census that 37.8% were black). The 1958 birth cohort was committing the majority of its delinquency from 1968 to 1978, a period of social movements and protests relating to civil rights, the women's movement, and the Vietnam peace movement. In addition, illegal drug use became an increasing problem in U.S. cities during this time, and the nature of crime changed to reflect this.

### The Racine Cohorts

Shannon (1988, 1991) studied three birth cohorts (1942, 1949, and 1955) from Racine, Wisconsin, a small midwestern industrial city. In order to maximize comparability to the London and Philadelphia cohorts, the analyses reported below pertain to the white or black male members of these three cohorts. This yields samples of 353 individuals from the 1942 cohort (4.2% black), 721 individuals from the 1949 cohort (6.1% black), and 1,067 from the 1955 cohort (9.9% black). Delinquent/criminal involvement in the Racine study was measured in almost the same way as in the Philadelphia study—by the number of *police contacts* for each year of age from ages 8–30 for the 1942 cohort, ages 8–25 for the 1949 cohort, and ages 8–22 for the 1955 cohort. Because the overall percentages of the Racine male cohorts with police contacts is very high due to inclu-

sion of reports for minor status offenses and traffic violations, we limit the analyses reported below to contacts for felonies/misdemeanors or the juvenile equivalents thereof.<sup>6</sup> This makes the Racine data more comparable to the London data. For the 1942 male cohort, 65.7% had at least one such official police contact by age 30, for the 1949 cohort, 64.8% had contact by age 25, and for the 1955 cohort, 55.8% had contact by age 22. For those with contacts, the average number of police contacts over ages 8–30 for the 1942 cohort was 5.48, for the 1949 cohort this average was 5.16 over ages 8–25, and for the 1955 cohort the average was 5.31 over ages 8–22.

This study's location in a relatively small city provides a distinct counterpoint to the previous two cohort studies from large cities. Racine's population was recorded at about 89,000 in the 1960 census, rose to about 95,000 in 1970, and then declined to about 86,000 in 1980. Of these totals, the percentage nonwhite was 5.4% in 1960, 11% in 1970, and 21.1% in 1980 (when it was separately estimated in the decennial census that 14.7% were black). While Shannon (1988, p. 51) expected to find "small, homogeneous" neighborhoods clearly delineated on the basis of measures such as race, socioeconomic status, and levels of crime in their ecological analyses, what he found instead was considerable overlap in neighborhoods and the delinquency/criminality of their occupants. In addition, during the cohort study there was an out-migration of both residents and businesses (e.g., grocery and liquor stores, gas stations, taverns, restaurants) from the inner city to newer, stable residential areas outside Racine.

## METHODS

Data from the London, Philadelphia, and Racine cohort studies are analyzed using semiparametric mixed Poisson regression models and their associated maximum-likelihood estimators with programs written in GAUSS (Aptech Systems 1992). Because prior research (Nagin and Land 1993; Land et al. 1996) suggests that a distinct age trajectory for the rate of offending is required for each point of support (latent class of offenders), these trajectories are allowed to vary for each class of offenders except for those classified as nonoffenders, whose age trajectory is fixed. Because the rates of offending across all ages are minimal for this group, lambda is held constant at .01 for all ages.

For each of the cohorts studied, it is necessary to specify the variables to be entered into the regression specifications (i.e., into the  $\mathbf{X}_i$  matrix for

<sup>6</sup> Inclusion of status offenses and traffic violations does not alter the main substantive conclusions of the analyses, but it does raise the estimated percentage of the cohort in the offender category of the 1955 cohort, for instance, to about 99.5%.

the model of eqq. [1a] and [1b] above) for the offending rate for individual  $i$  in period  $t$ ,  $\lambda_{it}$ . For the London cohort, in addition to age and age squared, a summary index, TOT <sub>$i$</sub> , of time-invariant personal characteristics of individual  $i$  (intelligence, parental child-rearing behavior, parental criminal activity, and a composite measure of risk taking and adventurous behavior) previously presented in Nagin and Farrington (1992a, 1992b) as covarying with the offending rate is included in the regression specification (see Nagin and Land 1993). In the case of the Philadelphia cohort, the regression specification for the  $\lambda_{it}$  includes age and age squared, race, and the socioeconomic status of the census tract of the individual's residence (as described above). For the Racine cohorts, the regression specification includes age, age squared, and race.<sup>7</sup>

Based upon previous work (Nagin and Land 1993; Land et al. 1996), a model that includes "intermittency," or periods of criminal activity possibly interspersed with periods of inactivity, is specified for all three cohort studies. The concept of intermittency generalizes the notion of criminal career onset and termination, structural elements that essentially divide the population into two classes: offenders and nonoffenders (Avi-Itzhak and Shinnar 1973). An intermittency parameter, denoted  $p_{it}$ , can account for periods during a usually active criminal career in which the rate of occurrence, or  $\lambda_{it}$ , is equal to zero. An individual in the sample is assumed to be in an active portion of his career during all time periods  $t$  with a probability  $p_{it}$  and considered to be inactive during periods with a probability  $1 - p_{it}$ . The addition of this intermittency parameter allows researchers to capture an important part of the theory behind the criminal careers paradigm: periods of inactivity may *not* be the sign of the termination of a delinquent or criminal career. Rather, they may only reflect a brief period of dormancy (Land et al. 1996). In the case of the London cohort, a separate regression equation is specified for the intermittency parameter  $p_{it}$  that includes its own constant term, TOT, a dichotomous variable  $Y_{t-1}$ , coded "1" if individual  $i$  had been convicted in the previous period and "0" otherwise, age and age squared to account for the possibility that intermittency is not homogeneous across time and persons (Nagin and Land 1993). In the case of the Philadelphia and Racine cohorts, a simpler regression specification for the intermittency parameter, using only  $Y_{t-1}$ , was found to be sufficient.

<sup>7</sup> While Shannon's (1988, 1991) studies also included a socioeconomic status index of Racine neighborhoods, this datum is missing for a number of observations. Therefore, in order to avoid sample selection biases that might arise from excluding observations due to missing data, we use only the race covariate in the analyses of the Racine cohorts.

TABLE 1  
BIC VALUES FOR THE LONDON AND PHILADELPHIA COHORTS

No. of Categories	London	Philadelphia	Philadelphia	Philadelphia
1 .....	3,073.59 (.13)	5,043.29 (.06)	9,542.87 (.06)	18,457.93 (.08)
2 .....	3,011.59 (.23)	4,946.18 (.13)	9,616.08 (.16)	17,987.09 (.16)
3 .....	2,962.94 (.32)	4,826.84 (.25)	9,417.23 (.26)	17,386.77 (.25)
4 .....	2,962.18 (.35)	4,822.88 (.26)	9,240.04 (.30)	17,268.73 (.29)
5 .....	HFTI	4,782.97 (.27)	9,186.92 (.34)	17,106.18 (.31)
6 .....	HFTI	HFTI	HFTI	HFTI
<i>N</i> .....	403	500	1,000	2,000

NOTE.—Pseudo- $R^2$  statistics are in parentheses. HFTI = Hessian matrix failed to invert.

## RESULTS

### Model Comparisons

Table 1 reports the computed BIC values for the London and Philadelphia cohorts with models as specified above and with numbers of latent classes of delinquent/criminal careers ranging from one (the conventional single-category Poisson regression model augmented with the intermittency parameter described above) to six.<sup>8</sup> In order to provide an intuitively interpretable measure of the goodness of fit of these Poisson and mixed Poisson regression models, the table also reports pseudo- $R^2$  values for each model. The pseudo- $R^2$  is computed as the square of the simple correlation between the observed  $y_{it}$  and the  $y_{it}$  predicted from each model. While there is no statistical theory that can justify use of the pseudo- $R^2$  as a statistic for testing the goodness of fit of Poisson-family models, the fact that we are using the same dependent variable and model specifications

<sup>8</sup> Because of space limitations and in order to focus attention on comparisons of the fit of alternative models and the nature of the delinquent/criminal careers they imply, we do not report the complete sets of all parameters estimated for the models described in this article. They are available from the authors on request. It should be noted, however, that the slopes remained consistent across models with varying numbers of categories. For example, with the Philadelphia cohort data, the constant term of the two category model is  $-0.629$ . For the three category model, the constant is  $-0.504$ , for the four category models it is  $-0.397$ , and for the five category model it is  $-0.191$ .



across models within cohorts permits use of the statistic for comparisons of the predictive accuracy across the models in table 1 (Land et al. 1996, p. 413). In addition to the BIC values and pseudo- $R^2$  statistics, we also report in table 1 the points in model complexity with respect to numbers of delinquent/criminal career categories at which our computer program could not compute standard errors of the regression coefficients (see n. 5).<sup>9</sup>

Consider first the results reported in the column in table 1 for the London cohort. Both the BIC values and pseudo- $R^2$  statistics in this column indicate a substantial improvement in the model specification and fit to the data as the number of classes of delinquent/criminal careers is increased from one to two.<sup>10</sup> There also are some improvements as the number of latent classes is increased from two to three. By contrast, the value of BIC decreases by only a fractional amount as the number of categories is further increased from three to four. In this case, the value of the BIC (as defined in appendix eq. [A3]) tells us that the decrease in the value of minus twice the log maximized likelihood in going from a three-category model to a four-category model—which adds four parameters (an adjustment to the overall intercept for the model plus coefficients for the age and age-squared terms and a coefficient for the sample density at the added point of support)—is barely larger than the penalty extracted by adding in 5.9989, the value of  $\log N$  when  $N = 403$ , multiplied by the total number of parameters of the model (18 in the four-category model). The pseudo- $R^2$  statistic also shows a more modest increase—from .32 to .35—in going from a three-category to a four-category than from a two-category to a three-category model. Nonetheless, application of the BIC rule implies that the model with four latent classes should be chosen.

Consider next the results for the three samples from the Philadelphia cohort reported in table 1. Recall that we estimated models for sample sizes of 500, 1,000, and 2,000 drawn randomly from the (white and black) male members of the Philadelphia cohort in order to maximize compara-

<sup>9</sup> For the London and all of the Philadelphia cohort samples, while our computer program converged for models with one more point of support than the maximum suggested as optimal by the BIC values in table 1, the Hessian matrix could not be inverted. As stated in n. 5, this implies that the additional point of support is not sufficiently distinct from the others for the inverse to exist. This condition also appears in the calculations for the Racine cohorts in table 2.

<sup>10</sup> The pseudo- $R^2$ s reported in table 1 for the conventional (single-category) Poisson regression models are of the same order of magnitude as those obtained by estimating the gamma-Poisson mixed or negative binomial model. In brief, while the negative binomial regression model provides improved estimates of standard errors (and therefore  $t$ -ratios) when the conditional means of the regressions do not equal the conditional variances, this generally does not improve the ability of the models to predict the values of the dependent variable (Land et al. 1996, p. 414).

TABLE 2  
BIC VALUES FOR THE RACINE COHORTS

No. of Categories	1942	1949	1955
1 .....	6,503.29 (.06)	11,492.92 (.06)	14,312.58 (.09)
2 .....	6,374.53 (.13)	11,350.09 (.14)	14,099.66 (.17)
3 .....	6,165.94 (.23)	10,555.93 (.28)	13,057.39 (.32)
4 .....	6,132.88 (.28)	10,449.05 (.32)	12,914.29 (.35)
5 .....	6,110.71 (.29)	10,459.07 (.33)	12,862.43 (.36)
6 .....	6,119.96 (.31)	...	HFTI
<i>N</i> .....	353	721	1,067

NOTE.—Pseudo- $R^2$  statistics are in parentheses. HFTI = Hessian matrix failed to invert.

bility to the London cohort and in order to address the issue of sensitivity of inferences about the number of delinquent/criminal career groupings or points of support of the mixing distribution to sample size. If the underlying mixing distribution of unobserved delinquent/criminal propensity is truly continuous, then statistical theory (Keifer and Wolfowitz 1956) implies that the inferred number of points of support will increase with sample size. On the other hand, if the mixing distribution is discrete or well approximated by a discrete distribution with a limited number of points of support, then the inferred value of this parameter should stabilize with increasing sample size. The values of BIC and the pseudo- $R^2$  statistics reported in the columns for the Philadelphia samples in table 1 suggest that the latter situation applies. Specifically, for all three samples, the BIC values imply that a semiparametric mixed Poisson model with five categories is optimal. Again, the pseudo- $R^2$ s show a substantial increase in going from a one- to a two-category model and from a two- to a three-category model. The pseudo- $R^2$ s also indicate that models with either four or five categories provide the best predictability of the dependent offense count variable.

Table 2 reports the BIC values and pseudo- $R^2$  statistics for the three Racine cohorts. These results show some variability in the optimal numbers of latent criminal career categories by cohort. According to the BIC values, a model with five categories provides an optimal fit to both the

1942 and the 1955 cohorts.<sup>11</sup> But the optimal model for the 1949 cohort is a four-category model. The BIC values also behave similarly to those in table 1—decreasing quite rapidly up through models with four latent classes and then less rapidly. The pseudo- $R^2$ s exhibit a similar pattern—relatively large increases up through models with four categories and then much more slowly.

The pseudo- $R^2$ s for the one-category Racine models in table 2 also show the impact of the single covariate, race of cohort member, available for inclusion as a covariate in the Racine cohorts. That is, the pseudo- $R^2$  values for these models are considerably smaller than those of the one-category model in table 1 for the London cohort—for which a richer set of measured delinquency/criminal propensity covariates is available. On the other hand, the Racine values are of comparable magnitude to those of the Philadelphia cohorts in table 1, which include both race and neighborhood socioeconomic status as covariates. Nonetheless, as the number of latent classes increases in the Philadelphia and Racine models, the pseudo- $R^2$ s increase in magnitude and become nearly as large as those of the London cohort. In brief, the incorporation of an optimal number of latent classes of delinquent/criminal offending purchases more predictive accuracy in the Philadelphia and Racine models than is the case for the London cohort.

In summary, across these cohort samples, drawn from populations in very different settings, the optimal number of latent classes is strikingly similar, four in London, five in Philadelphia, and four and five for the Racine cohorts.<sup>12</sup> We turn now to examining whether the offending trajectories of these latent groups are similar across samples.

### Age Curves of the Offending Classes

To assist in the substantive interpretation of the nature of the categories of delinquent/criminal careers optimally identified and estimated for each of the London, Philadelphia, and Racine cohorts, the age trajectories of expected (i.e., predicted on the basis of estimated model parameters) of-

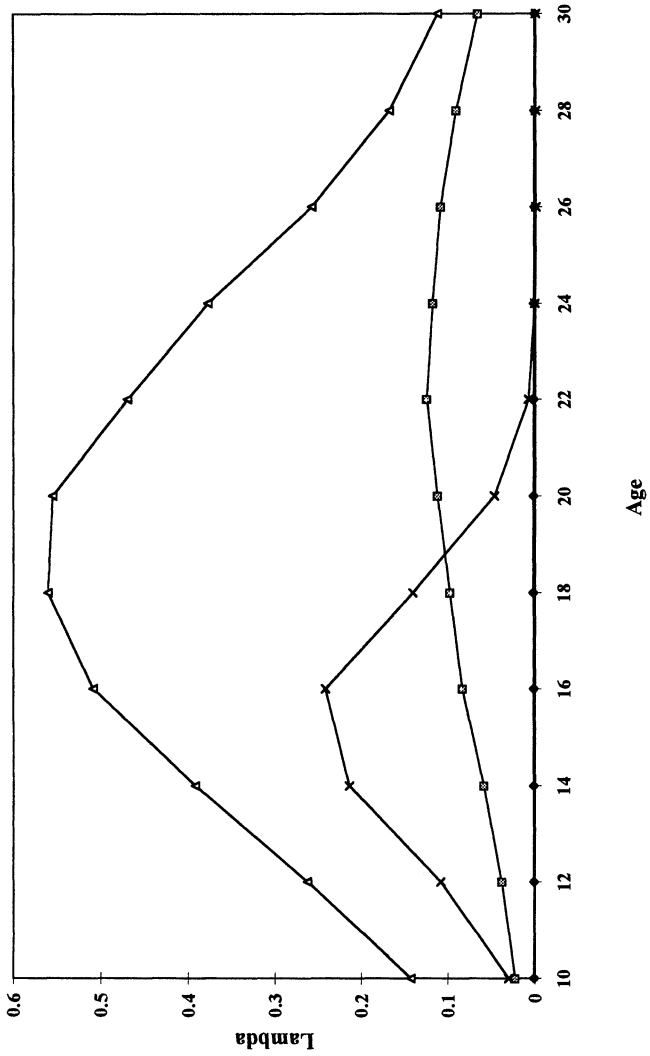
<sup>11</sup> A seven-category model for the 1942 cohort converged and yielded a BIC value smaller than the five-category model. However, the seven-category model also contained a number of redundant estimated parameters and therefore was not acceptable.

<sup>12</sup> As noted above, the regressor specifications used in the London, Philadelphia, and Racine mixed Poisson models differ substantially due to data limitations. The question then could be raised as to how robust our conclusions about optimal numbers of latent classes of offending trajectories are to these differing specifications. To address this question, we replicated the analyses reported in tables 1 and 2 as simple latent Poisson class models with no regressor variables. The results led to similar inferences about the optimal numbers of latent offending classes in all cohorts.

fending rates for each category implied by each model are plotted in figures 1–5 below. These graphs reveal a number of similarities as well as differences among the three cohorts. In the case of the London cohort, for example, the age trajectories of expected conviction rates in figure 1 clearly correspond to the four categories described in Nagin and Land (1993): the *nonoffenders* with virtually zero expected convictions across all ages; the *adolescence-peaked* grouping, which shows a rapid acceleration of expected conviction rates from age 10 through age 16 and then a rapid drop to a zero level by age 24; the *high-rate chronics*, who begin with a substantial conviction rate at age 10, reach a peak plateau of more than one conviction per two-year period at ages 18–20, and then experience a declining rate of convictions through the 20s (however, even by age 30, they still have a rate above zero); and, finally, the *low-rate chronics*, who similarly have long careers of offending (but at a rate that is much lower than the high-rate chronics) from age 10 through age 30 with essentially a plateau in the conviction rate in the ages 20 through 26 followed by a slow drop to age 30.

The age trajectories of expected police contact rates of the five-category model for the Philadelphia sample of 1,000 males are exhibited in figure 2. By comparison to the London cohort, in addition to a nonoffender trajectory, these graphs show *two adolescence-peaked trajectories*—one high and one low, and *two chronic trajectories*—one high and one low. In contrast to the London cohort, however, the expected offending rates for the high adolescence-peaked grouping, which rise to over three police contacts per year at ages 15–16, far exceed those of any of the other categories—including the high-rate chronics. Also, the graph for the high-rate chronics in figure 2 accelerates more slowly and peaks at about 0.9 police contacts expected per year at ages 17–19, which is two to three years beyond the peak ages for both adolescence-peaked groupings. This high-rate chronic trajectory, as for the London cohort, is characterized by a relatively slow rate of descent from the peak-rate ages. Also noteworthy in figure 2 is the age trajectory for the low adolescence-peaked category, which, like the high adolescence-peaked category, exhibits a high point at ages 15–16 but at a substantially lower annual rate of about 1 police contact per year. Finally, the low chronic offender category shows a trajectory in figure 2 that, like the high chronic category, peaks in the ages 17–19 but at a much lower rate of police contacts per year.

Figure 3 exhibits the age patterns of the expected rates of police contact for the optimal latent-class models for the 1942 Racine cohort. In addition to the nonoffenders, we find *one adolescence-peaked trajectory*—with very low rates that peak at about 0.25 felony/misdemeanor contacts per year at ages 16–17. The figure also shows *two chronic trajectories* similar to those found for the London cohort—*one high-rate* that plateaus at



◆ Nonoffenders —■— Low-Rate Chronic —▲— High-Rate Chronic —×— Adolescence-Limited

FIG. 1.—Predicted conviction rates by age for the London four-class model

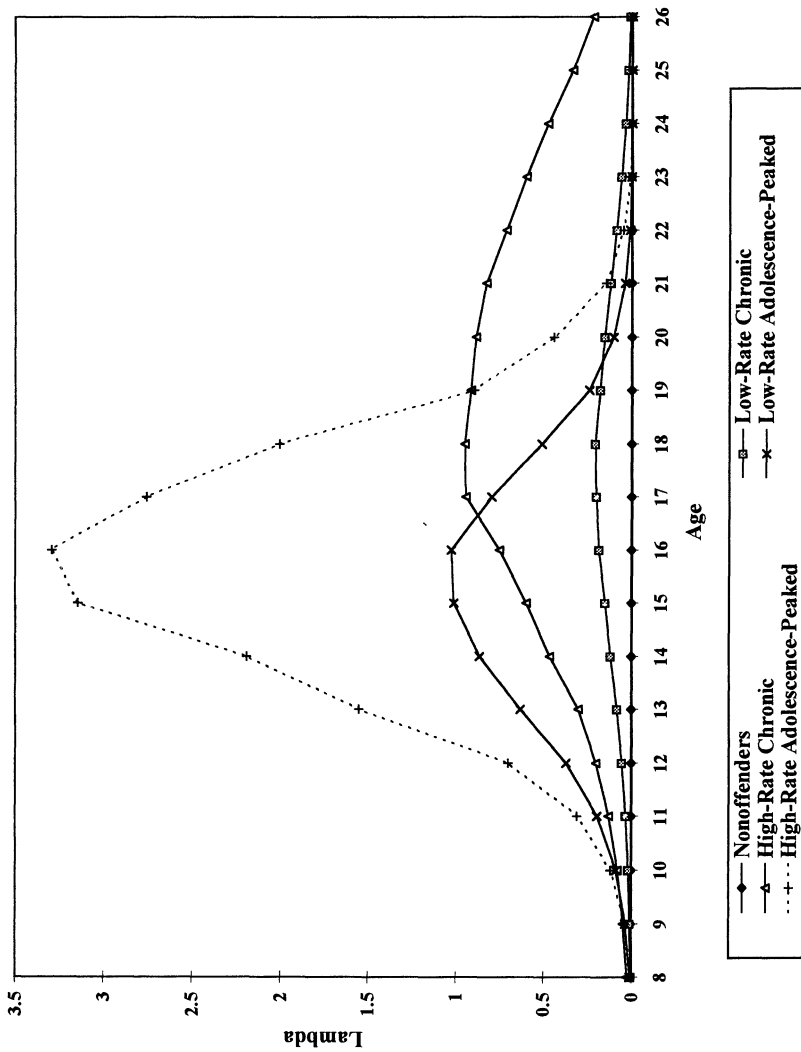


FIG. 2.—Predicted police contact rates by age for the Philadelphia five-class model

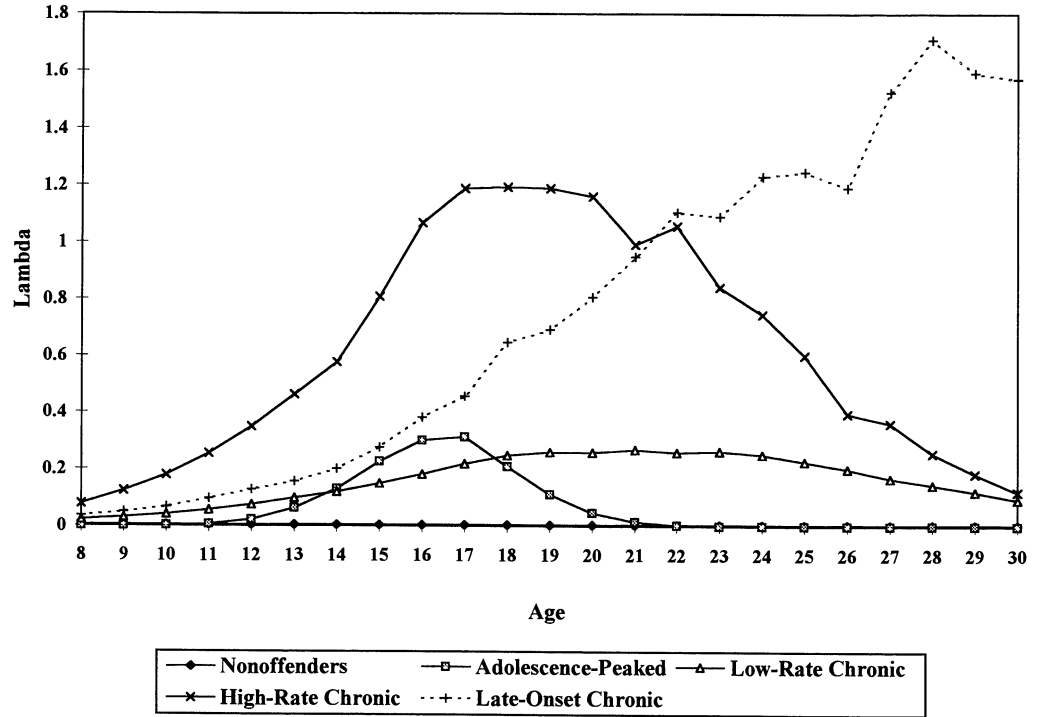


FIG. 3.—Predicted misemeanor/felony rates by age for the Racine 1942 five-class model

TABLE 3  
CHARACTERISTICS OF OFFENDING TRAJECTORIES IN THE LONDON, PHILADELPHIA,  
AND 1942 RACINE COHORTS

Cohort	Predicted Rate at Age 10	Predicted Rate at Age 26	Maximum Rate	Age at Maximum Rate
London:				
Adolescence peaked .....	.03	.00	.24	16
Low-rate chronic .....	.02	.12	.13	22
High-rate chronic .....	.14	.26	.56	18
Philadelphia:				
Low-rate adolescence peaked .....	.09	.00	1.02	16
High-rate adolescence peaked .....	.12	.00	3.29	16
Low-rate chronic .....	.02	.01	.21	18
High-rate chronic .....	.08	.22	.95	18
1942 Racine:				
Adolescence peaked .....	.00	.00	.31	17
Low-rate chronic .....	.04	.20	.26	23
High-rate chronic .....	.18	.40	1.19	18
Late-onset chronic .....	.07	1.19	1.71	28

about 1.2 contacts per year in ages 17–20 and *one low-rate* that reaches a plateau of 0.2–0.25 contacts per year for several years from age 18 through age 24. Last, in addition to these two chronic yet peaked trajectories, the 1942 cohort has *one late-onset chronic trajectory* that shows very little peaking behavior. It begins to accelerate at age 15 and continues to very high rates—increasing all the way to about 1.6 contacts per year at ages 28–30.

In summary, for the three cohorts for which we have offending data well beyond adolescence—London, Philadelphia, and 1942 Racine—the offense groupings are broadly similar but still differ in some important respects. Table 3 reports summary statistics on the offending trajectories displayed in figures 1, 2, and 3 for these cohorts—the rate at ages 10 and 26, the maximum rate, and the age at which this maximum is achieved. With one exception, in all three cohorts, the trajectories either rise quickly to a peak at about age 16 and then rapidly decline thereafter (adolescence-peaked offenders), or alternatively, rise until about age 18 and then begin a much more gradual decline. We call this later group chronic offenders because their rates are comparatively more stable. However, these broad classes are not in general homogeneous within or across cohort. In the Philadelphia sample we found evidence of at least two classes of adolescence-peaked offenders, what we call the high-rate and low-rate peaked



classes.<sup>13</sup> By contrast in the London and 1942 Racine samples we find evidence of only one adolescence-peaked group. Another possibly important difference across samples is that the highest rate of adolescent offending occurs among the high-rate chronics in the London and Racine samples but in the Philadelphia it is among the high adolescence peaked. Finally, in the 1942 Racine sample one grouping, the late-onset chronic class, follows an anomalous but also potentially important trajectory—a nearly steady rise throughout the observation period instead of a slow decline upon entering adulthood.

Figures 4 and 5 show the offense rate trajectories by age for the models of the 1949 and 1955 Racine cohorts. These trajectories must be interpreted with particular caution because data collection ended before the subjects had progressed very far into their adult years, age 25 for the 1949 sample and age 22 for the 1955 sample. Consequently, the estimated trajectories may not reliably capture downturns in offending that typically accompany the onset of adulthood. In the case of the four-class model of the 1949 cohort depicted in figure 4, *two* of the three nonzero offending age trajectories show *adolescence-peaked patterns*—one high and one low with each peaking at age 18 (at about 0.15 and 0.75 felony/misdemeanors per year) and declining quite rapidly by the early 20s. This cohort also has *one high-rate chronic age trajectory* similar to that of the London cohort—with a plateau of greater than two felonies/misdemeanors per year in the ages 17 through 23. From the five-category model of the 1955 cohort displayed in figure 5, in addition to *one low-rate chronic pattern*, we find *two adolescence-peaked trajectories*—one an *early onset pattern* that begins to accelerate at ages 8–10, peaks at about three felony/misdemeanor contacts per year at ages 15–17 and then declines to low levels by the early 20s, and one *late-onset pattern* that begins to accelerate at ages 14–15 and peaks at less than one contact per year at ages 17–20. The figure also shows *one high-rate chronic trajectory*—similar to the high-rate chronics in the London cohort—that begins to accelerate at ages 8–10, attains a plateau of over one contact per year over ages 14–19, and then declines relatively slowly.

### Characteristics of the Latent Offending Classes

After estimation of the optimal mixed Poisson models of criminal careers identified above for each cohort, individual sample members can be sorted

<sup>13</sup> Because the group that we labeled low-rate chronics has a negligible rate offending by age 26 they are also arguably another distinctive group of adolescence-peaked offenders. But this class possesses very little, if any, peaking of its offending rate during adolescence; hence, we term it a chronic pattern.

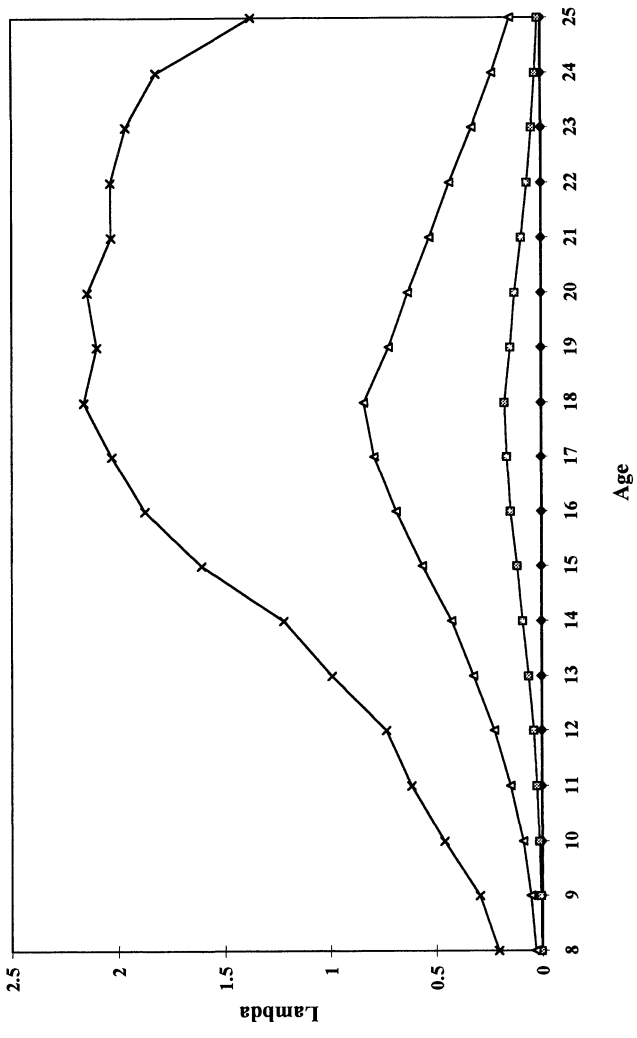


FIG. 4.—Predicted misdemeanor/felony rates by age for the Racine 1949 four-class model

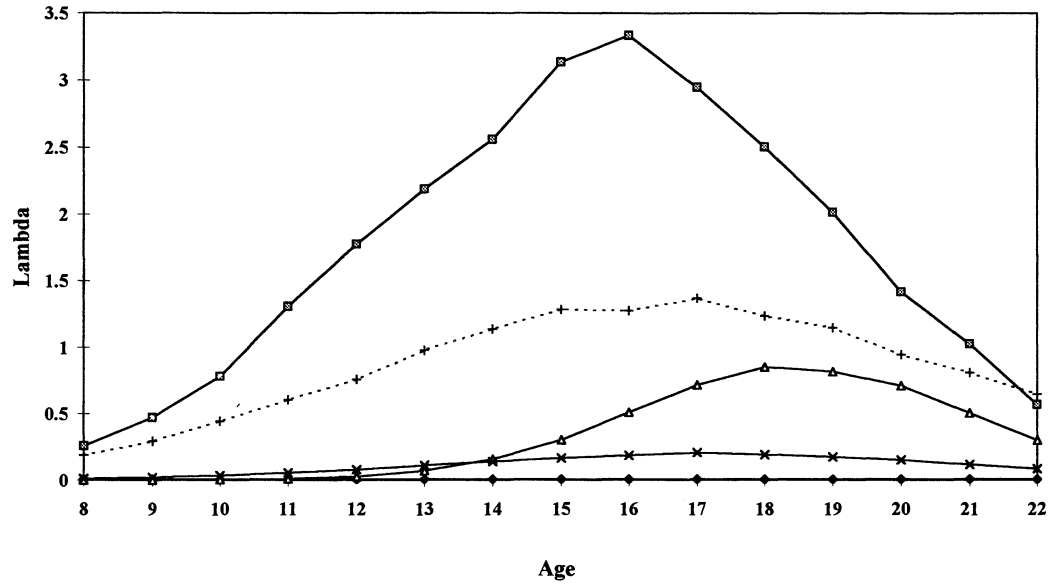


FIG. 5.—Predicted misdemeanor/felony rates by age for the Racine 1955 five-class model

TABLE 4  
OFFENDING CLASSES IN THE LONDON AND PHILADELPHIA COHORTS

Category	Cohort Members (%)	Cohort Offenders (%)	N of Offenses	Average N of Offenses	Cohort Offenses (%)
London:					
Nonoffenders .....	64.0	. . .	0	0	0
Adolescence peaked .....	12.7	35.2	103	2.02	16.1
Low-rate chronics .....	9.9	27.6	90	2.25	14.1
High-rate chronics .....	13.4	37.2	447	8.28	69.8
Philadelphia:					
Nonoffenders .....	60.8	. . .	1	.002	.0005
Low-rate adolescence peaked .....	8.6	21.9	509	5.9	27.31
High-rate adolescence peaked .....	1.0	2.6	183	18.3	9.82
Low-rate chronics .....	21.3	54.3	390	1.8	20.92
High-rate chronics .....	8.3	21.2	781	9.4	41.90

using the following “maximum probability” procedure: Based on the model coefficient estimates, the probability of observing each individual’s longitudinal pattern of offending is computed conditional on his being, respectively, in each of the latent classes. The individual then is assigned to the grouping with the highest probability. Based on this procedure, for example, an individual with a cluster of offenses in adolescence but not thereafter will, in all likelihood, be assigned to (one of) the adolescence-peaked classes. Note that the maximum probability sorting rule does not guarantee perfect assignments. Based on this rule, for instance, individuals with no recorded offenses are always categorized as nonoffenders. Some such individuals may in fact be members of one of the active offending groups but simply had no *recorded* offenses.

For the London and Philadelphia cohorts, table 4 reports the following summary statistics computed after sample members are sorted into the criminal career classes identified: the first column is the percentage of all cohort members in the class, the second column is the percentage of cohort offenders in the class, the third column is the total number of offenses committed by individuals in the class, the fourth column is the average number of offenses committed per class member, and the last column is the percentage of cohort offenses committed by individuals in the class. For the London cohort, these statistics show that the highest per person rate of offending (across all ages 10–30 observed) belongs to the high-rate chronics (8.28 convictions per person) followed by the low-rate chronics (2.25 convictions) and the adolescence limiteds (2.02 convictions). The high-rate chronics also account for about 70% of total convictions despite making up only 13.4% of the total cohort and 37.2% of the offenders.

TABLE 5  
OFFENDING CLASSES IN THE RACINE COHORTS

Category	Cohort Members (%)	Cohort Offenders (%)	N of Offenses	Average N of Offenses	Cohort Offenses (%)
1942:					
Nonoffenders .....	34.6	. . .	1	.008	.08
Adolescence peaked .....	20.1	30.7	140	2.0	11.0
Low-rate chronics .....	31.4	48.1	416	3.7	32.7
High-rate chronics .....	8.8	13.4	419	13.5	32.9
Late-onset chronics .....	5.1	7.8	296	16.4	23.3
1949:					
Nonoffenders .....	35.4	. . .	1	.004	.04
Low-rate adolescence peaked .....	39.8	61.6	482	1.7	20
High-rate adolescence peaked .....	19.4	30.0	976	7.0	40.5
High-rate chronics .....	5.4	8.4	952	24.4	39.5
1955:					
Nonoffenders .....	44.5	. . .	3	.006	.09
Early onset adolescence peaked .....	2.2	4.1	595	24.8	18.8
Late-onset adolescence peaked .....	15.4	27.7	826	5.0	26.1
Low-rate chronics .....	30.1	54.2	624	1.9	19.8
High-rate chronics .....	7.8	14.0	1,111	13.4	35.2

In the case of the Philadelphia cohort, it can be seen from table 3 that the highest per person rate of offending (across all ages 8–26 in the panel) belongs to the high adolescence-peaked class (18.3 police contacts per person) followed by the high-rate chronics (9.41 contacts), the low adolescence-peaked (5.92 contacts), and the low-rate chronics (1.83 contacts). Of the higher offending-rate classes, the high-rate chronics make up about 8% of the total cohort, about 21% of the offenders, and account for about 42% of the cohort offenses. By comparison, the high adolescence-peaked class makes up only 1% of the cohort, about 2.6% of the offenders, and yet accounts for about 10% of all cohort offenses. The low-rate chronics also are interesting: They make up a larger grouping—about 21% of the cohort and over 50% of the offenders—but account for only about 21% of all cohort offenses.

Summary statistics for the three Racine cohorts are shown in table 5. A first point to note about these cohorts is that the percentages in the nonoffenders category—about 35% in the 1942 and 1949 cohorts and about 45% in the 1955 cohort—are much smaller than in either the London (64%) or Philadelphia (60.8%) cohorts. This is true even though the definition of the dependent offending variable in the Racine cohorts—felonies or misdemeanors or their juvenile equivalents—is more serious

than in Philadelphia (all police contacts) and less serious than in London (convictions). Second, consistent with a closer community social control/policing context in Racine, the average numbers of offenses committed per member of the high-rate chronic categories in the three cohorts—13.5 (1942), 24.4 (1949), and 13.4 (1955)—are higher than those in London or Philadelphia. Third, the percentages of the cohorts in these high-rate chronic categories in Racine—8.8% in 1942 (13.9% if we add in the late-onset chronic in this cohort), 5.4% (1949), and 7.8% (1955)—generally are closer to that found above in Philadelphia (8.3%) than to that found in London (13.4%). Fourth, the percentages of the Racine cohorts in the adolescence-peaked categories—20.1% (1942), 59.2% (1949), and 17.6% (1955)—are higher than in London (12.7%) or Philadelphia (9.6%). Fifth, of all the cohorts studied, it bears repeating that only the 1942 Racine cohort evidences the late-onset chronic offender grouping—making up 5.1% (18 individuals) of the 353 cohort members, having the highest average number of offenses committed per category member (16.4) of the cohort, and accounting for nearly one fourth of all the cohort's offenses.

## DISCUSSION

Applications of semiparametric mixed Poisson regression models have been important in the development of micromodels of delinquent/criminal careers (Land et al. 1996; Land and Nagin 1996; Nagin and Land 1993), yet the question of how many latent classes of criminal careers should be included in these models has not been systematically addressed in the work. The empirical results presented in this article suggest that there is no simple answer to this question but still, under a wide range of circumstances, four to five groups appear sufficient. Notwithstanding, the shapes of the offending trajectories over age are likely to depend heavily on the type of dependent variable used to count offenses (e.g., self-report data, arrests, convictions), on the nature of the cohort studied, and on the community context from which it was drawn.

In the case of convictions recorded for members of the working-class cohort studied in the large city of London by West and Farrington (1973), we found that a four-category model fit the data best. Moreover, this model possesses a high-rate chronic category of offenders who make up about 13% of the total cohort and account for about 70% of the total number of convictions of the cohort. By contrast, for a large city such as Philadelphia and the comprehensive citywide 1958 cohort study of police contacts for status offenses, delinquencies, misdemeanors, and felonies by Tracy et al. (1990), we found that a five-category model—which splits both the chronic and the adolescence-peaked categories of offenders into low and high groups—yields a best fit to the data. In this model, the high

adolescence-peaked category actually shows the highest average number of offenses per category member but accounts for less than 10% of the total cohort offenses. By comparison, the high-rate chronics in Philadelphia compose about 8% of the cohort and account for about 42% of the total cohort offenses. In total, the two chronic offending categories in Philadelphia make up nearly 30% of the total cohort and account for about 63% of the total cohort offenses, whereas the two adolescence-peaked categories in Philadelphia make up less than 10% of the total cohort and account for about 37% of the total cohort offenses.

Finally, for the relatively small city of Racine and the study of police contacts for delinquencies, felonies, and misdemeanors in the 1942, 1949, and 1955 cohorts by Shannon (1988, 1991), we found total numbers of latent offending classes—five, four, and five, respectively—that are consistent with those found in London and Philadelphia, although the precise age trajectories differ somewhat. For instance, while the percentages of the Racine cohorts in the high-rate chronic categories are about the same as in Philadelphia and lower than in London, they account for smaller percentages of the offenses committed by the cohort members. In part, this is due to the larger percentages of the Racine cohorts who have one or more offenses than in the other two cities. This relatively high prevalence of offending may be, in part, a function of more active surveillance/social control activities on the part of citizens and police in this relatively small city.<sup>14</sup> The Racine cohorts also show larger percentages in the adolescence-peaked categories, which suggests that most young males in Racine in these years experienced a period of offending during their adolescence/young adult years but that very few continued to offend at a high rate beyond their early 20s.

In analyses of the three samples from the Philadelphia cohort of sizes 500, 1,000, and 2,000, we also found that a model with five categories of offenders was robust to sample size. This is a remarkable finding. It implies that, rather than merely representing a discrete approximation to an underlying continuous distribution of unobserved delinquent/criminal propensity, the small number of latent offending categories estimated in our models may represent distinct classifications of cohort offenders with respect to age trajectories of offending that are meaningful in and of themselves. Without replication and extension beyond the present study, this

<sup>14</sup> Lyle W. Shannon, in a personal communication to Kenneth C. Land, August 20, 1996, notes a strong emphasis on record keeping in the Racine police department (e.g., offenses which may have gone unrecorded in other cities were carefully recorded in Racine). According to Shannon, college-educated police were put in charge of the record keeping in Racine, which increased its accuracy. He suggests that this made it much easier to monitor known chronic offenders in Racine than in a larger city, such as Philadelphia.

point should not be overinterpreted—leading to a reification of the identified classes of offenders. Some research evidence suggests, however, that similar latent classes can be discerned using self-report data. Research (Laub, Nagin, and Sampson 1998) that uses longitudinal interview data from Glueck and Glueck's (1950) sample of 500 boys finds that four latent classes of offenders exist based on the self-report data. Research (McDermott and Nagin 1998) using National Youth Survey data (Elliott et al. 1983) also suggests the presence of multiple groups of offenders with varying age-crime trajectories, each affected differently by covariates included in the models. This indicates that the significance of our findings may hold true across varying sample sizes as well as types of dependent variables (e.g., convictions, arrests, contact with police, self-reported offending).

With these findings in hand, some general propositions can be generated to guide subsequent studies of criminal careers. First, despite the fact that the specific number of classes that can be identified will depend on (a) the type of dependent variable used, (b) the nature of the sample on which the cohort study is based, and (c) the characteristics of the community from which it is drawn, it is likely that one or both of two general age patterns of offending will be found: (1) an *adolescence-peaked pattern* for which the offense rate rises rapidly to a peak between ages 15 and 18 but which drops rapidly toward a near zero level by about age 22, and (2) a *chronic offender pattern*, which does not peak until ages 17–21 and drops much more slowly through the 20s. Second, with a comprehensive cohort sample across the entire population of a large city, it is possible, as in the Philadelphia models estimated here, that these two patterns will *bifurcate into high and low offending groupings* (and possibly into an even larger number of groupings). It also is possible that the trajectories may exhibit either *early or late onset* patterns.<sup>15</sup> But the essential trajectory character-

<sup>15</sup> Indeed, the latent class space of offending trajectories identified by the analyses reported here appears to have three dimensions: chronic vs. adolescence peaked, high rate vs. low rate, and early onset vs. late onset. Using these dichotomies for the dimensions yields a latent-class space with  $2^3 = 8$  cells. All of the trajectories identified herein can be classified into one of these cells. But none of these offending trajectories exhibit patterns that fit into two cells—the early-onset, low-rate, adolescence-peaked trajectory and the late-onset, high-rate, adolescence-peaked trajectory. This is not to say, of course, that it is empirically impossible for such trajectories to occur. Evidently, however, a low-rate, adolescence-peaked offending trajectory is not likely to have an early onset, and a high-rate, adolescence-peaked trajectory is not likely to have a late onset. Thus, the empirically effective space of latent classes of offending trajectories consists of the six remaining cells: (1) late-onset, high-rate chronics; (2) early-onset, high-rate chronics; (3) late-onset, low-rate chronics; (4) early-onset, low-rate chronics; (5) early-onset, high-rate adolescence peakeds; and (6) late-onset, low-rate, adolescence peakeds. Future empirical research and theoretical developments should explicate the developmental dynamics and institutional interactions that make more likely the empirical presence of offending trajectories in these six cells of the latent class space.



istics of adolescence-peaked versus chronic offending should be detectable, regardless of the number of groupings found. While these two basic categories of offending patterns are similar to the “life-course persistents” and “adolescence limiteds” identified by Moffitt (1993) with respect to antisocial behavior, the question as to whether the developmental mechanisms and traits postulated by Moffitt or those cited by Cohen and Vila (1996), or possibly others, differentiate the groupings requires research that extends beyond this article.

These findings also speak to the conjectures put forth by Cohen and Vila (1996). First, our findings indicate that criminological analysts and theorists do indeed need to recognize the existence of at least a second class of offending career pattern—the adolescence-peaked pattern—in addition to the high-rate chronic pattern. Second, using our estimates from the comprehensive cohort studies in Philadelphia and Racine, we conjecture that the high-rate chronics may make up 5%–10% of the population—a larger percentage than conventional estimates of the male sociopath population. Thus, it is likely that this category of offender is drawn from marginally as well as strongly antisocial individuals as well as possibly other individuals who do not exhibit antisocial personality characteristics but who, for other reasons, develop a chronic offending pattern. Third, while our Philadelphia-based estimates of the size of the high-rate chronics suggests that they account for only about 21% of the offender population, if we combine both the high- and low-rate chronics, we arrive at an estimated 76% of the offender population that accounts for about 63% of the cohort offenses—estimates that are near the upper bound suggested by Cohen and Vila (1996). Fourth, again using the Philadelphia-based estimates, the adolescence-peaked categories appear to make up about 10% of the population and to exhibit the kinds of offending patterns that possibly could be explained by the “life-course variability in bonding” theory of Sampson and Laub (1993). But, whether comparisons are limited only to the high categories of the chronic and adolescence-peaked patterns or extended to the combined high and low groupings for each type of trajectory, the chronics make up at least as large a portion of the offender population in London and Philadelphia than do the adolescence peakeds. This also is true for two of the three Racine cohorts. This runs counter to Moffitt’s (1993) hypothesis about the relative sizes of the two groups.

When viewed from a life-course perspective, these findings have some important implications for how we study crime, as well as how we study sociology. We must recognize the possibility that more hidden heterogeneity exists in social science data than previously acknowledged. The effects of age, cohort or sample composition, and historical setting all play important roles in influencing individual development, hence the variation in trajectories of behavior over time. Social context must be viewed as a

“force in development” (Elder and O’Rand 1995), which has the power to alter trajectories of myriad types of behavior—be they economic activities, publication careers, sexual behavior, or participation in crime. Semi-parametric mixed Poisson regression models make it possible to uncover such heterogeneity, which may allow researchers to better get at the causal mechanisms behind a myriad of behaviors. The finding of multiple categories where only one was assumed—whether it be in offending or sexual behavior—may challenge not only the theories sociologists develop, but also how they are empirically tested.

In conclusion, the empirical findings reported have a number of implications for future theorizing and research on delinquent/criminal careers. To begin with, rather than continuing to theorize about delinquent/criminal offending in general, our findings imply that theories of the etiology of delinquent/criminal offending should explicitly recognize the existence of two or more qualitatively different classes with respect to age patterns of offending. Moreover, such theories should seek to identify characteristics of the distinct groupings that support either persistence and even increasing offending rates across the adolescent ages into adulthood, or a peak in adolescence and subsequent decline. Characteristics that distinguish high-rate from low-rate trajectories as well as early onset from late onset also should be theorized. To evaluate hypotheses about such distinguishing characteristics, additional empirical research then is needed along the lines developed in Nagin and Land (1993) and Nagin, Farrington, and Moffitt (1995) with the London cohort.

## APPENDIX

### Statistical Decision Rules for an Optimal Number of Categories of Criminal Careers

A *two-stage procedure* for estimation and comparison of models was suggested in Nagin and Land (1993): first-stage models include the addition of a single parameter (adjustment to the constant term as illustrated in eqq. [1a] and [1b] above) to the regression model for each additional point of support, while second-stage models incorporate more flexibility in the parametric specifications. Because the first stage of the model building process is intended to provide baseline models for decisions about the optimal number of classes, it includes only a single age trajectory of offending. This age trajectory usually is specified as quadratic, with age and age-squared terms (incorporated into the  $X_{it}$  matrix in the model of eqq. [1a] and [1b] above). If the age trajectory of offending is, in fact, linear in age, or constant across all ages for a particular cohort data set, then one or both of the regression coefficients for the age and age-squared terms should fail

to reach statistical significance. In second-stage models, the age trajectories of offending are allowed to vary across groups, thus generating more accurate parameter estimates for comparison of offender categories.

First-stage models yield likelihood functions that can be used to generate likelihood-ratio test statistics, allowing for the comparison of models containing differing numbers of points of support. Taking minus twice the difference of the log likelihoods yields the conventional *likelihood-ratio test statistic*. In large samples for models that satisfy certain “regularity” conditions (Cox and Hinkley 1974, p. 281), this test statistic is approximately distributed as a chi-square variable, with degrees of freedom equal to the difference of numbers of parameters between the two models. Two problems in the application of this standard testing procedure to decisions about the number of points of support in mixed Poisson regression models were presented in Land et al. (1996). First, it is generally the case that for mixtures of parametric distributions with two or more parameters, the likelihood-ratio testing approach fails because there is not a unique way of obtaining the null hypothesis from the alternative hypothesis (Titterington, Smith, and Makov 1985, pp. 4–5). In the case of the second-stage models presented in Nagin and Land (1993), for instance, one can obtain a two point-of-support model from a three point-of-support model either by setting, say, the  $m_3$  parameter for the proportion of the population at the third point of support to zero or by setting the three parameters defining the quadratic age trajectory of the third point of support equal to those for one of the other points of support. What, then, are the appropriate degrees of freedom for the chi-squared distribution of the likelihood-ratio test—one or three? This problem can be solved by application of the likelihood-ratio test to first-stage models, wherein a reduction in the number of points of support from  $n$  to  $n - 1$  results in a change of only one degree of freedom (i.e., either by setting the proportion  $m_3$  to zero or by setting the single parameter in the regression specification for the third group equal to the value for another group; see Land et al. 1996, p. 425). Even in this case, however, problems remain in the application of the likelihood-ratio test (Land et al. 1996, n. 24), for setting a point of support to zero places a constraint on the boundary of the parameter space—which violates one of the standard regularity conditions on which asymptotic distribution theory is based (Titterington et al. 1985, p. 156). Thus, even for first-stage models, there is no assurance from asymptotic statistical theory that the likelihood-ratio test will have a chi-squared distribution.

Given these problems with the application of the likelihood-ratio test statistic to the question of how many points of support, it appears that other decision rules should be explored. In the analyses presented here, we apply the so-called *Bayesian information criterion (BIC)*. BIC is equal

to minus twice the Schwarz criterion, a statistic derived by Schwarz (1978) for the purpose of estimating the dimension of a statistical model. BIC recently has been related to the concept of Bayes factors by Kass and Raftery (1995) and Raftery (1995a). For the comparison of two models  $M_2$  and  $M_1$ , the *Bayes factor* for  $M_2$  against  $M_1$  is defined as the ratio:

$$B_{21} = p(D/M_2)/p(D/M_1), \quad (\text{A1})$$

where  $p(D/M_k)$  denotes the probability of the data given model  $M_k$ . This probability can be computed as the integrated likelihood of model  $M_k$ , so that the Bayes factor is the ratio of the integrated likelihoods of the two models. When the analyst regards the models  $M_1$  and  $M_2$  as equally probable a priori (i.e., before model parameters are computed from observed data) so that  $p(M_1) = p(M_2) = .5$ , the Bayes factor  $B_{21}$  is equal to the posterior odds in favor of  $M_2$ .

Kass and Raftery (1995) and Raftery (1995a) argue that BIC can be used to approximate the Bayes factor for comparisons of nested or un-nested models under a fairly general set of conditions according to the formula:

$$\text{BIC}_{21} \approx -2 \log B_{21}. \quad (\text{A2})$$

Moreover, the BIC for a model can be computed according to the simple formula:

$$\begin{aligned} \text{BIC} = & -2(\log \text{maximized likelihood of the model}) \\ & + (\log N)(\text{no. of parameters in the model}), \end{aligned} \quad (\text{A3})$$

where  $N$  denotes sample size. In the usual situation where prior information is small or highly uncertain relative to the information provided by the data, the model with the highest posterior probability is the model that minimizes BIC computed in this way (Kass and Raftery 1995, p. 790). Note that this statistical criterion favors model parsimony by extracting a penalty for complicating a model (by adding parameters) that increases with the log of the sample size. Furthermore, this *BIC (or Schwarz) criterion for model selection* embodies the intuitive notion that, when the analyst complicates a model by adding parameters, the payoff in terms of a decrease in the log maximized-likelihood function of the model should be larger than this penalty.<sup>16</sup>

<sup>16</sup> It should be noted that the use of BIC both as an approximation to Bayes factors and as a statistical decision rule for model uncertainty is not without its critics; see the exchange between Gelman and Rubin (1995) and Raftery (1995b). Our judgment, based on the theoretical arguments as well as the empirical applications presented in Kass and Raftery (1995) and Raftery (1995a), is that the BIC approximation and decision rule for model selection work well.

REFERENCES

- Abbott, Andrew, and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Data." *American Journal of Sociology* 96:144–85.
- Aptech Systems. 1992. *GAUSS 3.0 Systems and Applications Manual*. Maple Valley, Wash.: Aptech Systems.
- Avi-Itzhak, Benjamin, and Reuel Shinnar. 1973. "Quantitative Models in Crime Control." *Journal of Criminal Justice* 1:185–217.
- Birkelund, Gunn Elisabeth, Leo A. Goodman, and David Rose. 1996. "The Latent Structure of Job Characteristics of Men and Women." *American Journal of Sociology* 102:80–113.
- Blumstein, Alfred, Jacqueline Cohen, Jeffrey A. Roth, and Christy A. Visher, eds. 1986. *Criminal Careers and "Career Criminals,"* 2 vols. Washington, D.C.: National Academy Press.
- Cameron, A. Colin, and Pravin K. Trivedi. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1:29–53.
- Caspi, Avshalom, and Daryl J. Bem. 1990. "Personality Continuity and Change across the Life Course." In *Handbook of Personality: Theory and Research*, edited by Lawrence A. Pervin. New York: Guilford.
- Caspi, Avshalom, and Terrie E. Moffitt. 1993. "When Do Individual Differences Matter? A Paradoxical Theory of Personality Coherence." *Psychological Inquiry* 4:247–71.
- Clogg, Clifford C. 1994. "Latent Class Models." Pp. 311–59 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by G. Arminger, C. C. Clogg, and M. E. Sobel. New York: Plenum.
- Cohen, Lawrence E., and Bryan J. Vila. 1996. "Self-Control and Social-Control: An Exposition of the Gottfredson-Hirschi/Sampson-Laub Debate." *Studies on Crime and Crime Prevention* 5:125–50.
- Cox, D. R., and D. V. Hinkley. 1974. *Theoretical Statistics*. London: Chapman & Hall.
- Elder, Glen H. Jr. 1975. "Age Differentiation and the Life Course." *Annual Review of Sociology* 1:165–90.
- . 1992. "Life Course." Pp. 1120–30 in *Encyclopedia of Sociology*, vol. 3. New York: MacMillan.
- Elder, Glen H., Jr., and Angela M. O'Rand. 1995. "Adult Lives in a Changing Society." Pp. 452–75 in *Sociological Perspectives in Social Psychology*, edited by Karen S. Cook, Gary Alan Fine, and James S. House. Boston: Allyn & Bacon.
- Elliott, Delbert, et al. 1983. *The Prevalence and Incidence of Delinquent Behavior: 1976–1980*. Boulder, Colo.: Behavioral Research Institute.
- Gelman, Andrew, and Donald E. Rubin. 1995. "Avoiding Model Selection in Bayesian Social Research." *Sociological Methodology* 25:165–74.
- Glueck, Sheldon, and Eleanor Glueck. 1950. *Unraveling Juvenile Delinquency*. New York: Commonwealth Fund.
- Goodman, Leo A. 1974a. "The Analysis of Qualitative Variables When Some of the Variables Are Unobservable. Part I—a Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–1259.
- . 1974b. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.
- . 1987. "New Methods for Analyzing the Intrinsic Character of Qualitative Variables Using Cross-Classified Data." *American Journal of Sociology* 93:529–83.
- Gottfredson, Michael, and Travis Hirschi. 1990. *A General Theory of Crime*. Palo Alto, Calif.: Stanford University Press.

- Greenberg, David F. 1991. "Modeling Criminal Careers." *Criminology* 30:149–55.
- Hirschi, Travis, and Michael R. Gottfredson. 1995. "Control Theory and the Life-Course Perspective." *Studies on Crime and Crime Prevention* 4:131–42.
- Kass, Robert E., and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90:773–95.
- Keifer, Jack, and Jacob Wolfowitz. 1956. "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters." *Annals of Mathematical Statistics* 27:887–906.
- Land, Kenneth C., Patricia L. McCall, and Daniel S. Nagin. 1996. "A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models with Empirical Applications to Criminal Careers Research." *Sociological Methods and Research* 24:387–442.
- Land, Kenneth C., and Daniel S. Nagin. 1996. "Micromodels of Criminal Careers: A Synthesis of the Criminal Careers and Life Course Approaches via Semiparametric Mixed Poisson Regression Models, with Empirical Models." *Journal of Quantitative Criminology* 12:163–91.
- Laub, John H., Daniel S. Nagin, and Robert J. Sampson. 1998. "Trajectories of Change in Criminal Offending: Good Marriages and the Desistance Process." *American Sociological Review*, in press.
- McDermott, Shaun, and Daniel Nagin. 1998. "Same or Different? Comparing Offender Groups and Covariates over Time." In *Handbook of Law and Social Science: Youth and Justice*, edited by S. O. White. New York: Plenum, in press.
- Moffitt, Terrie E. 1993. "'Life-Course-Persistent' and 'Adolescent-Limited' Antisocial Behavior: A Developmental Taxonomy." *Psychological Review* 100:674–701.
- Nagin, Daniel S., and David P. Farrington. 1992a. "The Onset and Persistence of Offending." *Criminology* 30:501–23.
- . 1992b. "The Stability of Criminal Potential from Childhood to Adulthood." *Criminology* 30:235–60.
- Nagin, Daniel S., David P. Farrington, and Terrie E. Moffitt. 1995. "Life-Course Trajectories of Different Types of Offenders." *Criminology* 33:111–40.
- Nagin, Daniel S., and Kenneth C. Land. 1993. "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model." *Criminology* 31:327–62.
- Nagin, Daniel S., and Raymond Paternoster. 1991. "On the Relationship of Past to Future Delinquency." *Criminology* 29:163–89.
- Oakes, David. 1988. "Semi-Parametric Models." Pp. 367–69 in *Encyclopedia of Statistical Sciences*, vol. 8. Edited by S. Kotz and N. L. Johnson. New York: Wiley.
- Raftery, Adrian E. 1995a. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–64.
- . 1995b. "Rejoinder: Model Selection Is Unavoidable in Social Research." *Sociological Methodology* 25:185–96.
- Rowe, David C., D. Wayne Osgood, and W. Alan Nicewander. 1990. "A Latent Trait Approach to Unifying Criminal Careers." *Criminology* 28:237–70.
- Sampson, Robert J., and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points through Life*. Cambridge, Mass.: Harvard University Press.
- . 1995. "Understanding Variability in Lives through Time: Contributions of Life-Course Criminology." *Studies on Crime and Crime Prevention* 4:143–58.
- Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461–64.
- Shannon, Lyle W. 1988. *Criminal Career Continuity: Its Social Context*. New York: Human Sciences Press.
- . 1991. *Changing Patterns of Delinquency and Crime: A Longitudinal Study in Racine*. Boulder, Colo.: Westview.

American Journal of Sociology

- Titterton, D. M., A. F. M. Smith, and U. E. Makov. 1985. *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- Tracy, Paul E., Marvin E. Wolfgang, and Robert M. Figlio. 1990. *Delinquency in Two Birth Cohorts*. New York: Plenum.
- West, Donald J., and David P. Farrington. 1973. *Who Becomes Delinquent?* London: Heinemann.
- . 1977. *The Delinquent Way of Life*. London: Heinemann.
- Wolfgang, Marvin E., Robert M. Figlio, and Thorsten Sellin. 1972. *Delinquency in a Birth Cohort*. Chicago: University of Chicago Press.