Sociological Methods & Research

http://smr.sagepub.com

On the Development of Different Kinds of Criminal Activity

ROBERT BRAME, EDWARD P. MULVEY and ALEX R. PIQUERO

Sociological Methods Research 2001; 29; 319
DOI: 10.1177/0049124101029003003

Research has established that individuals who participate in one form of antisocial behavior are at relatively high risk of participating in other forms of antisocial behavior as well. In this article, the authors rely on recent methodological advances to study the longitudinal development of two different forms of antisocial behavior: serious violent offenses and other offenses. The results of their analysis of two Philadelphia birth cohorts suggest that trajectories of offending differ more in degree than in kind and that relative standing in the distribution of violent offending is associated with relative standing in the distribution of other types of offending activity.

On the Development of Different Kinds of Criminal Activity

ROBERT BRAME
University of Maryland
EDWARD P. MULVEY
University of Pittsburgh School of Medicine
ALEX R. PIQUERO
Northeastern University

1. INTRODUCTION

Investigators of antisocial behavior and criminal activity struggle regularly with the question of whether a general theory is applicable to all kinds of antisocial behavior or whether different theories are needed to explain the occurrence of different types of antisocial behavior. This struggle is evidenced by recent debates over the appropriateness of theorizing about a risk behavior syndrome (Jessor 1992; see also Huesmann et al. 1984) within the field of developmental psychopathology or the value of developing a general theory of crime (Gottfredson and Hirschi 1990) within the field of criminology. The basic question is whether involvement in different kinds of antisocial

AUTHORS' NOTE: *The authors' names are listed in alphabetical order. We would like to thank Alfred Blumstein, Daniel Nagin, and Larry Wasserman for helpful discussions related to this article. The authors are also affiliated with the National Consortium on Violence Research.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 29 No. 3, February 2001 319-341

acts is indicative of a general tendency toward antisocial behavior or whether distinct etiological processes drive different types of antisocial behaviors.

There is a great deal of evidence that different types of antisocial behaviors co-occur. For example, using exploratory factor analytic techniques, Donovan and Jessor (1985) found that a single underlying factor accounted for the intercorrelations among a set of problem behaviors in both male and female high school and college students, in two, independent waves of longitudinal data. In a replication study, Donovan, Jessor, and Costa (1988) used data from a sample of 11th- and 12th-grade students surveyed in 1985 to further examine the generality-of-deviance hypothesis. Their results indicated that a single factor reproduced the correlations between alcohol use, marijuana use, sexual intercourse, and general deviant behavior. Other researchers (see, e.g., Gilmore et al. 1991; Osgood et al. 1988) have detected multiple factors, rather than a single factor, but they still come to the conclusion that, given the positive correlations among the factors, deviance may be a single factor at a higher order of analysis.

In addition, numerous researchers (Moffitt 1993; Farrington 1986) have shown that the prevalence of different antisocial behaviors shifts over the period of adolescence. It is well documented that crime rates aggregated across the population tend to be highest among older adolescents and young adults (Blumstein et al. 1986; Hirschi and Gottfredson 1983). Other forms of antisocial behavior (e.g., drug use, lying) also show consistent developmental trends (Loeber and LeBlanc 1990; Elliott, Huizinga, and Menard 1989). Some theorists have explained these trends as indicating a robust tendency for all individuals to approach their peak levels of antisocial activity during late adolescence (Gottfredson and Hirschi 1990). Others have proposed that there are at least two (and possibly multiple) subpopulations of adolescents, the great proportion of whom commit antisocial acts during adolescence and others who start young and persist in these behaviors over the life span (Loeber 1982; Moffitt 1993; Patterson and Yoerger 1993; Nagin and Land 1993).

In this article, we assemble evidence that should inform discussion of these issues and promote research to address them more fully. Our efforts draw on recent work by Nagin and his colleagues (Nagin and Land 1993; Land, McCall, and Nagin 1996; Land and Nagin 1996;

Laub, Nagin, and Sampson 1998; Nagin and Tremblay 1999; Roeder, Lynch, and Nagin 1999; Nagin 1999), who have been involved in a program of research that provides a very powerful set of tools (based on finite mixture or latent class models) for studying the population distribution of changes in criminal activity over the life span. Using methods developed within this framework, we examine how involvement in two different forms of criminal activity changes as individuals age. Our primary objective is to see whether variation in the age curve for the two types of offenses under study is due primarily to differences between individuals in their general tendency to offend or, alternatively, whether different groups exhibit fundamentally different developmental patterns for each type of criminal activity. The approach used in this article allows us to assess whether, and to what extent, there is an association between developmental trends associated with different forms of criminal or antisocial activity.

Although the methodology developed by Nagin and colleagues is relatively new, we believe it provides several distinct advantages over other, alternative methods for investigating this set of problems. The primary advantage of this approach is that it allows us to estimate a statistical model whose parameters govern the joint longitudinal distribution of (1) serious violent and (2) other forms of criminal activity. Such a model has several useful features.

First, it does not require that we build the mixture from any specific probability distribution. We are free to choose any probability distribution that makes sense for our specific problem. This differs from many approaches that force researchers to work with, for example, the normal distribution even when the outcomes being studied are clearly not normally distributed.

Second, because each individual is observed at multiple time periods, it is unlikely that offense counts at different time periods are independent of each other. Simple forms of repeated cross-sectional analysis are quite restrictive in this sense. A more realistic model, such as the one used in this article, will allow for this kind of within-subject dependence.

Third, prior research shows that, at least for some offenders, the incidence of criminal activity within individuals changes as they grow older (Blumstein et al. 1986; Loeber and LeBlanc 1990; Farrington 1986; Moffitt 1993; Hirschi and Gottfredson 1983). It may be the case that very different parameters govern the growth of offending in different subpopulations.

While many forms of growth-curve modeling assume that growth-curve parameters are drawn from a multivariate normal distribution, the approach used here assumes that these parameters are drawn from a multinomial distribution whose shape is determined completely by the data. In this sense, the models used in this article can be viewed as semiparametric rather than fully parametric (see discussion in Nagin and Tremblay 1999; Land et al. 1996).

The article is organized as follows. In section 2, we describe the two data sets used in this analysis. Section 3 provides details about our statistical model. The analysis results are presented in section 4, and section 5 presents our conclusions.

2. DATA

In this article, we examine longitudinal data sets from two general population Philadelphia birth cohorts (Wolfgang, Figlio, and Sellin 1972; Tracy, Wolfgang, and Figlio 1990). The Philadelphia birth cohorts are composed of longitudinal data that retrospectively identify individuals born in the same city during the same year and follow them for a number of years. Both advantages and disadvantages of such data have been described in detail elsewhere (Tracy et al. 1990; Farrington et al. 1990).

The 1945 Philadelphia birth cohort study was the first large-scale study of its kind undertaken in the United States. It systematically investigated the juvenile police records of all 9,945 males who were born in the city of Philadelphia in 1945 and who resided in the city between ages 10 and 18 (Wolfgang et al. 1972). One of the most well-known findings arising from this study was that 6 percent of the cohort was responsible for 52 percent of all the police contact generated by the cohort through age 17.

In similar fashion, the 1958 Philadelphia birth cohort study investigated the juvenile police records of the 13,160 males who were born in Philadelphia in 1958 and who resided in the city between ages 10 and 18. Although the 1958 cohort exhibited significantly more serious, violent criminal activity than the 1945 cohort, many of the other findings from the 1945 study were replicated in the later study.

Several limitations of the two Philadelphia birth cohort studies should be recognized at the outset. First, the data cover only one city (Philadelphia). This, of course, affects our ability to generalize the results to other cities. Second, the analysis considers only the juvenile criminal activity of males. Therefore, the results do not have anything to say about female juvenile criminal activity. Third, the indicators of criminal involvement come only from arrest reports. The archived versions of the Philadelphia birth cohort data do not contain any information for self-reported involvement in criminal activity. (For problems associated with this issue, see Moffitt 1993; Nagin and Land 1993.) Fourth, we can only be certain that all members of the cohort resided in the city through age 17. Although the original investigators have compiled criminal history information through age 26 for the 1958 cohort, this information is based only on local (not state) records (Tracy and Kempf-Leonard 1996). Therefore, in this analysis, we are unable to say anything about adult criminal activity in either cohort. Fifth, both cohorts are limited in that they do not contain any information on whether individuals were incarcerated at any point during their lifetimes. To the extent that it occurred, "time off the street" could influence the conclusions we draw from this analysis (see Piquero et al. 2001).

In both data sets, we have defined serious violent criminal activity as any police contact (Philadelphia) for one of the following offenses: homicide, forcible rape, robbery, and aggravated assault. Police contacts for all other kinds of criminal activity are placed in the "other" category. We have arranged the data so that the number of police contacts for each type of criminal activity (serious violent or other) is recorded for each individual at each year from age 8 to age 17.

3. STATISTICAL MODEL

In this section, we describe our methods in detail. First, assume that we observe criminal offending histories for each of $i = 1, 2, \ldots, N$ individuals. Furthermore, assume that each individual's offending history is composed of a set of variables, y_i, that captures the number of offenses associated with that individual at each of the $t = 1, 2, \ldots, T$ discrete periods of time under study. We denote the number of offenses associated with a given individual, i, at time, t, by y_{it}.

Now, let us assume that the multivariate distribution of the y_{it}s can be adequately represented by a finite mixture of $j = 1, 2, \ldots, K$ processes. Sometimes, the K components of the mixture are referred to as latent classes (Everitt and Hand 1981:3). We will adopt that convention in this article. Within each of the latent classes associated with the mixture, we further assume that the process that generates y_{is} is independent of the process that generates y_{it} where $s \neq t$. Clogg (1995) refers to this as the "local independence assumption," and it is a common (but not necessary) feature of latent class models.

It is important to note that local independence is different from global independence. Under local independence, we only assume that elements of the multivariate distribution are independent conditional on the latent class. Under global independence, we would have to assume that elements of the multivariate outcome distribution are unconditionally independent. Thus, the assumption of local independence (which we make) is much weaker than the assumption of global independence (which we do not make).

Since y_{it} takes the form of a count of discrete events (e.g., crimes or arrests), we assume that, conditional on the latent class, j, the outcomes arrive according to a Poisson process. The probability mass function associated with a Poisson process where the random variable, y_{it}, has mean λ_{it} is given by

$$p(y_{it} = n \mid \lambda_{it}) = \frac{\exp(-\lambda_{it})\lambda_{it}^{n}}{n!}, \tag{1}$$

where $n = 0, 1, 2, \ldots$ and subject to the usual constraint that $\lambda_{it} > 0$ (Johnson, Kotz, and Kemp 1992).

In the special case where there are no latent classes, we could obtain the probability distribution for the $1 \times T$ vector, y_i, by the product of T independent Poissons:

$$p(y_i = n \mid \lambda) = \prod_{t=1}^{T} \left(\frac{\exp(-\lambda_{it})\lambda_{it}^{n}}{n!} \right), \tag{2}$$

where again n can take on positive integer values. Since λ_{it} is free to vary across time periods, we can allow it to depend on time by invoking the following parameterization:

$$\lambda_{it} = \exp(\alpha + \beta_t t) \tag{3}$$

so that λ_{it} is simply an exponential function of t, and α and β_t are estimated from the data using maximum likelihood methods. In this article, we adopt a more flexible functional form that allows for log-quadratic dependence of t on λ_{it}:

$$\lambda_{it} = \exp(\alpha + \beta_t t + \beta_{t^2} t^2), \tag{4}$$

where α, β_t, and β_{t^2} are all estimated from the data using maximum likelihood methods.

In practice, the case where there are no latent classes will often be too simplistic when working with restrictive distributions like the Poisson. To relax this constraint, we condition λ_{it} on the latent class, j. Formally, we have

$$\lambda_{it|j} = \exp(\alpha_j + \beta_{t|j} t + \beta_{t^2|j}\, t^2), \tag{5}$$

where all of the parameters contributing to $\lambda_{it|j}$ are allowed to vary across each of the $j = 1, 2, \ldots, K$ latent classes. The unconditional probability mass function for the multivariate distribution of y_i when there is more than one latent class (i.e., $K \geq 2$) is given by

$$p(y_i = n|\lambda, \pi) = \sum_{j=1}^{k} \pi_j \left[\prod_{t=1}^{T} \left(\frac{\exp(-\lambda_{it|j})\lambda_{it|j}^{n}}{n!} \right) \right], \tag{6}$$

where π_j is the estimated probability that an individual drawn at random from the population under study is a member of latent class j. Since the classes are, in theory, mutually exclusive and exhaustive, it follows that $\sum_{j=1}^{K} \pi_j = 1.0$. In general, it is necessary to estimate $K - 1$ of the elements of $\pi = \{\pi_{j=1}, \pi_{j=2}, \ldots, \pi_{j=K}\}$ by the method of maximum likelihood, and the final element is then determined by the difference between 1.0 and the sum of the estimated elements (Everitt and Hand 1981:97-98). The parameter estimates associated with each of the K

latent classes are then treated as random draws from a multinomial distribution with cell probabilities given by the vector, π. The log-likelihood function associated with this model is given by

$$\log_e(L) = \sum_{i=1}^{n} \log_e \left(\sum_{j=1}^{K} \pi_j \left[\prod_{t=1}^{T} \left(\frac{\exp(-\lambda_{it|j})\lambda_{it|j}^{y_{it}}}{y_{it}!} \right) \right] \right), \tag{7}$$

where $\log_e(\bullet)$denotes the natural logarithm or the logarithm to the base e.

The model as described thus far is the version that has been used in most applied work. Recent arguments presented by Roeder et al. (1999), Nagin and Tremblay (1999), and Nagin (1999), however, illustrate a variety of ways in which the above estimator can be extended. As suggested by these articles, one obvious way to extend this model is by studying the longitudinal distribution of two different categories of criminal activity: violent and nonviolent offenses. We now turn our attention to the implementation of this more complex model.

First, let q_{it} denote the number of violent criminal acts associated with individual i at time t. Next, let r_{it} denote the number of "other" criminal acts associated with individual i at time t. In the special case where q_{it} and r_{it} are independent, we can write their joint probability distribution as the product of two independent Poissons:

$$p(q_{it} = n_q, r_{it} = n_r | \theta_{qit}, \theta_{rit}) = \left(\frac{\exp(-\theta_{qit})\theta_{qit}^{n_q}}{n_q!} \right) \times \left(\frac{\exp(-\theta_{rit})\theta_{rit}^{n_r}}{n_r!} \right), \tag{8}$$

where $n_q = 0, 1, 2, \ldots, n_r = 0, 1, 2, \ldots$, and the Poisson parameters are θ_{qit} and θ_{rit}, respectively.

Next, we parameterize the mean functions for violent and other criminal activity, respectively. The mean function for violent criminal activity is given by

$$\theta_{qit} = \exp(\gamma_{q0} + \gamma_{qt}t + \gamma_{qt^2}\ t^2), \tag{9}$$

and the mean function for other kinds of criminal activity is given by

$$\theta_{rit} = \exp(\gamma_{r0} + \gamma_{rt} t + \gamma_{qr^2}\ t^2), \tag{10}$$

where the parameter vector γ is estimated from the data using maximum likelihood methods. These equations are valid if there is only one latent class. As discussed above, however, it is useful to relax this constraint. Thus, for violent criminal activity, we have the conditional mean

$$\theta_{qit|j} = \exp(\gamma_{q0|j} + \gamma_{qt|j} t + \gamma_{qt^2|j} t^2), \tag{11}$$

and for other kinds criminal activity, the conditional mean is given by

$$\theta_{rit|j} = \exp(\gamma_{r0|j} + \gamma_{rt|j} t + \gamma_{rt^2|j} t^2), \tag{12}$$

for each of the $j = 1, 2, \ldots, K$ latent classes. An important feature of this parameterization is that membership in class j associated with violent criminal activity implies membership in class j associated with other kinds of criminal activity. The analytical basis for this result is fully developed in Roeder et al. (1999) (see also Nagin 1999).

We now invoke the assumption of local independence for violent and other kinds of criminal activity. Under this assumption, the probability mass function for the pair (q_{it}, r_{it}) is given by

$$p(q_{it}, r_{it} | \theta_{qit|j}, \theta_{rit|j}) = \left(\frac{\exp(-\theta_{qit|j}) \theta_{qit|j}^{n_{qit}}}{n_{qit}!} \right) \times \left(\frac{\exp(-\theta_{rit|j}) \theta_{rit|j}^{n_{rit}}}{n_{rit}!} \right). \tag{13}$$

Conditional on the latent class, j, we now have the assumption that q_{it} and r_{it} are independent of each other. As we discussed above in the context of within-subject dependence that accrues from repeated observations, the local independence assumption does not logically imply anything at all about the marginal relationship between q_{it} and r_{it}. Combining this framework with the repeated observation model above, we have the joint probability mass function:

$$p(q_i, r_i | \theta_{qi|j}, \theta_{ri|j}) = \prod_{t=1}^{T} \left[\left(\frac{\exp(-\theta_{qit|j}) \theta_{qit|j}^{n_{qit}}}{n_{qit}!} \right) \times \left(\frac{\exp(-\theta_{rit|j}) \theta_{rit|j}^{n_{rit}}}{n_{rit}!} \right) \right]. \tag{14}$$

This result is crucial because it allows us to investigate the expected time path of violent and other forms of criminal activity conditional on the latent class. To assign this interpretation to the results, however, we have to maximize the log-likelihood function over both outcomes simultaneously. Within the latent class framework, the full log-likelihood function is given by

$$\log(L) = \sum_{i=1}^{n} \log\left[\sum_{j=1}^{K} \pi_j \left(\prod_{t=1}^{T}\left[\left(\frac{\exp(-\theta_{qit|j})\theta_{qit|j}^{q_{it}}}{q_{it}!}\right)\times\left(\frac{\exp(-\theta_{rit|j})\theta_{rit|j}^{r_{it}}}{r_{it}!}\right)\right]\right)\right], \tag{15}$$

and once again, we impose the constraint that $\sum_{j=1}^{K} \pi_j = 1.0$.

So long as K is known (Roeder et al. 1999; Nagin 1999), maximization of this log-likelihood function can be achieved with standard tools such as Newton-Raphson (or other forms of numerical optimization) or expectation-maximization (EM) algorithms (see, e.g., Tanner 1996; Dempster, Laird, and Rubin 1977). As discussed by Nagin (1999), an important practical problem is how to optimize the choice of K. A common approach is to evaluate the Schwarz (1978) criterion, also referred to as the Bayesian information criterion (Kass and Raftery 1995:778; Kass and Wasserman 1995:928; Wasserman 1997), which can be viewed as an approximation to the natural logarithm of the Bayes factor, $B_{1,2}$, when we place equal prior probability mass on models 1 and 2, respectively. It is calculated by

$$\log(B_{1,2}) \approx \log(L_{M_1}) - \log(L_{M_2}) + \left(\frac{d_2 - d_1}{2}\right)\log(n), \tag{16}$$

where M_j denotes model j, d_j is the dimension of M_j, and n is the number of observations.[1] In general, the Bayes factor is interpreted as the ratio of one posterior model probability to another. Specifically, we have

$$B_{1,2} = \frac{p(M_1 \mid q_i, r_i)}{p(M_2 \mid q_i, r_i)}. \tag{17}$$

Guidelines for interpreting the magnitude of Bayes factors are well documented in the statistics literature (see, e.g., Kass and Raftery 1995; Wasserman 1997; Nagin 1999). In general, if $B_{1,2} < 0.1$ (i.e., M_2

is at least 10 times more probable than M_1 after conditioning on the available data), we have strong evidence for M_2, and if $B_{1,2} > 10$ (i.e., M_1 is at least 10 times more probable than M_2), we have strong evidence for M_1.

A central issue in the interpretation of any model is whether (and to what extent) the model's predictions correspond to what is actually observed in the data. Recent articles by Land et al. (1996) and Laub et al. (1998) provide details about how such a post hoc assessment can be conducted. We will adopt the approach described by these authors in our investigation.

To implement this particular post hoc procedure, it is necessary to first obtain a vector that is composed of the posterior probability of latent class membership for each of the $j = 1, 2, \ldots K$ latent classes. To obtain the posterior probability of membership in latent class j, we use Bayes's theorem, which gives

$$p(i \in j | q_i, r_i, \theta_j, \pi_j) = \frac{p(q_i, r_i | \theta_j) \times \pi_j}{\sum_{j=1}^{K} p(q_i, r_i | \theta_j) \times \pi_j}, \tag{18}$$

where π_j is the estimated unconditional probability of membership in latent class j, $p(q_i, r_i|\theta_j)$ is the likelihood function conditional on the event $i \in j$, and the denominator is a normalizing constant in any given application. If we conduct this calculation for each of the K latent classes for each individual, then we obtain the vector of posterior latent class membership probabilities for each individual.

With this vector in hand, it is now possible to identify the latent class to which each individual has the highest posterior probability of belonging. If we are willing to treat latent class membership as if it were observed (rather than latent), we can see whether (and to what extent) the observed offending patterns for individuals who have been classified into a particular latent class correspond to the expected offending patterns for that latent class. Ideally, the observed offending patterns for individuals assigned to a given class will correspond closely to the offending patterns that we expect to see in that class on the basis of the statistical model. In the next section, we will consider both the expected and actual longitudinal patterns of violent and other kinds of offending activity associated with each latent class.

4. RESULTS

For each data set, we present the parameter estimates associated with the statistical models described in the previous section. Since these models contain numerous nonlinear terms, however, they are somewhat difficult to interpret. In light of this, we use graphical methods to investigate the growth curves in serious violent activity and other kinds of offending that are implied by these models.

In Figure 1, we display average frequencies of serious violent and other forms of criminal activity by age within the 1945 Philadelphia birth cohort data set. Figure 2 presents the same information for the 1958 data set. This is a very simple analysis insofar as the graphs in Figures 1 and 2 simply show how the average number of violent and other kinds of offenses changes over time for the population as a whole. An important theme that emerges from these figures is that the age-graded patterns of violent activity and other kinds of criminal activity seem to have roughly the same shape, although violent activity is responsible for only a small proportion of the total amount of criminal activity at every age within each cohort.

An interesting question not answered by the graphs in Figures 1 and 2 is whether there is an association between violent and other kinds of offending over time. An advantage of the methods that we described in the previous section is that we will be able to answer this question, and in addition, we will be able to describe the nature of the relationship.

Table 1 presents the parameter estimates associated with the age-graded development of violent and other forms of criminal activity for the 1945 and 1958 Philadelphia birth cohorts. The term γ_{01} denotes the intercept associated with the first latent class for violent offending (top half of the table) and other forms of criminal offending (bottom half of the table). Similarly, γ_{02} and γ_{03} denote the intercepts associated with the second and third latent classes, respectively.

Also associated with each latent class is a pair of coefficients that captures the change in offending frequency as a function of age. Consider the first latent class. The first term in the pair is γ_{1t}, which measures the amount of change in the expected natural logarithm of the offending frequency associated with a unit change in age. The second term, γ_{1t}^2, measures the amount of change in the expected natural logarithm of offending frequency associated with a unit change in the square

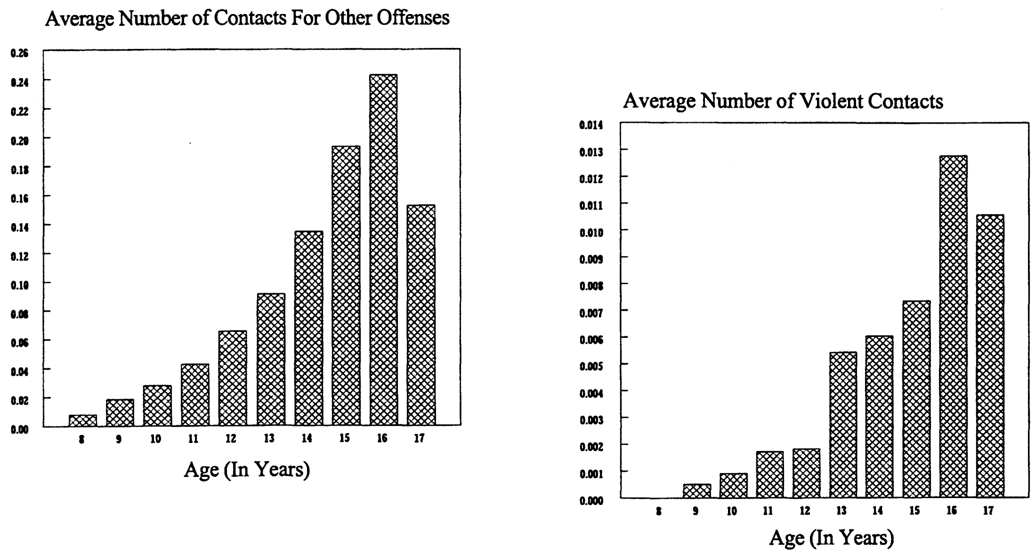


Figure 1: **Overall Violent and Other Offense Trajectories in 1945 Philadelphia Birth Cohort (N = 9,944)**

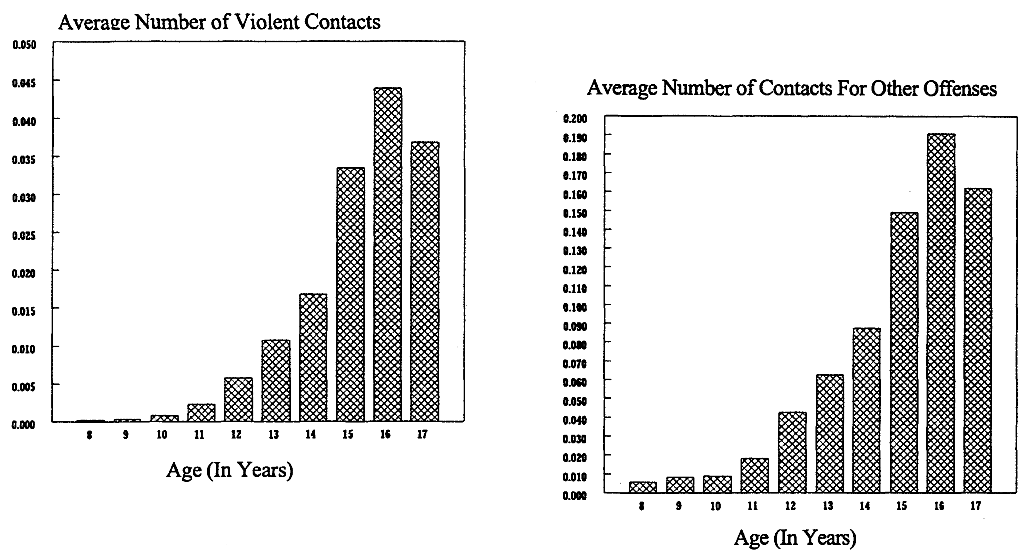


Figure 2: Overall Violent and Other Offense Trajectories in 1958 Philadelphia Birth Cohort (N = 13,160)

TABLE 1: Parameter Estimates for 1945 and 1958 Philadelphia Birth Cohort Models

	1945 Cohort (N = *9,944)*		*1958 Cohort* (N = *13,160)*	
Parameter	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*
Serious violent model				
γ_{01}	–16.099	8.618	–29.174	5.088
$\gamma_1(t)$	0.871	1.263	2.971	0.718
$\gamma_1(t^2)$	–0.195	0.455	–0.938	0.251
γ_{02}	–18.276	3.738	–23.251	2.295
$\gamma_2(t)$	1.646	0.530	2.381	0.315
$\gamma_2(t^2)$	–0.448	0.186	–0.656	0.107
γ_{03}	–18.373	3.153	–31.133	2.798
$\gamma_3(t)$	2.019	0.455	3.880	0.387
$\gamma_3(t^2)$	–0.618	0.162	–1.224	0.133
Model for other offenses				
γ_{01}	–13.278	0.987	–12.819	1.157
$\gamma_1(t)$	1.113	0.144	0.895	0.168
$\gamma_1(t^2)$	–0.295	0.052	–0.191	0.060
γ_{02}	–19.684	1.166	–19.292	1.055
$\gamma_2(t)$	2.354	0.160	2.174	0.145
$\gamma_2(t^2)$	–0.717	0.054	–0.621	0.050
γ_{03}	–10.973	0.619	–13.393	0.762
$\gamma_3(t)$	1.552	0.095	1.778	0.113
$\gamma_3(t^2)$	–0.526	0.036	–0.565	0.041
Mixing proportions				
π_1	.813	.007	.832	.005
π_2	.157	.007	.147	.004
π_3	.030	—	.021	—
Log-likelihood	–28282.94		–33737.89	

of age divided by 10. This division ensures that the linear and quadratic terms have roughly the same magnitude and is helpful in stabilizing the optimization routine (but it has no substantive impact on the results).

By including a term that captures the square of age, we are able to allow explicitly for curvilinear growth in criminal activity over the life span. The mixing proportions at the bottom of the table convey the unconditional probability that an individual drawn at random from the population under study comes from each latent class. In a model with three latent classes, only two of these proportions have to be estimated because the third will always be a linear function of the other two (i.e., $\pi_3 = 1 - \pi_1 - \pi_2$).

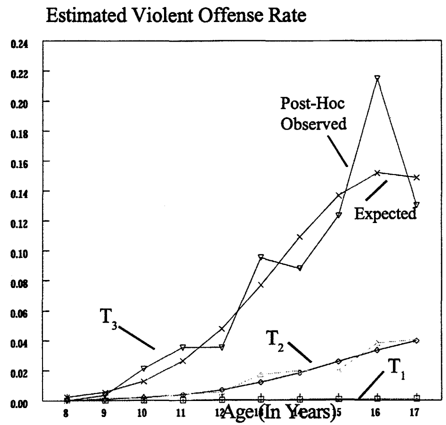
Using the Schwarz criterion described above, we concluded that a model with $K = 3$ latent classes was a significant improvement on the model with $K = 2$ latent classes in each cohort. Since we were unable to attain convergence with a model that had $K = 4$ latent classes, the model with $K = 3$ latent classes yields the optimal choice for K. As Nagin and Land (1993) show, it is possible for the mixture to reach a saturation point. In these situations, the latent classes simply collapse on one another. In all likelihood, our inability to attain convergence with a model allowing for $K = 4$ latent classes can be attributed to reaching this saturation point.

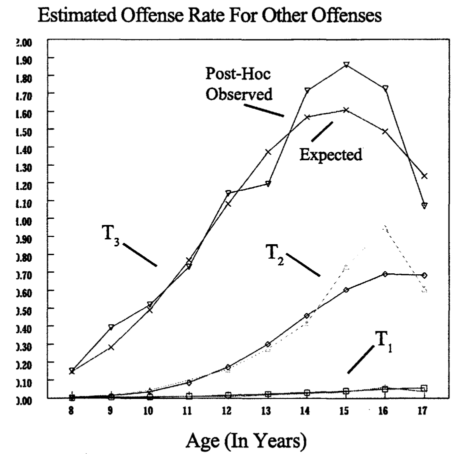
An important substantive feature of the parameter estimates in Table 1 is the sign pattern for the linear and the quadratic age terms. All of the linear terms are positive, while all of the quadratic terms are negative. This implies an initial positive relationship between age and offense frequency followed by a period of leveling-off or even a period of decreasing offense frequency (see also Nagin and Land 1993).

As mentioned above, the numerous nonlinear terms in these models complicate the interpretation of the parameter estimates. To simplify the interpretation, we have constructed graphs that show the expected average offense frequency (e.g., the offense rate) at each age for each of the three latent classes. We refer to the time path of average crime frequency for a particular latent class as a *trajectory*.

Figures 3 and 4 show the violent and "other" offense trajectories for each of the three latent classes in the 1945 and 1958 Philadelphia data sets. An important aspect of these figures is that people who occupy the first trajectory, T_1, in their violent offending also occupy the first trajectory in other forms of offending activity. Thus, the people in T_1 in the violent graphs are the same persons who are in T_1 in the graphs for other kinds of offenses. The same is true for T_2 and T_3. Both figures also suggest quite reasonable agreement between the expected population trajectories and those trajectories that are observed after individuals have been classified into the latent class to which they have the highest probability of belonging.

Several conclusions can be drawn from these two figures. First, there is considerable fluctuation in violent and nonviolent activity during the period spanning ages 8 to 17. In all three trajectories, we observe a substantial increase in both categories of offense frequency over the adolescent period. In the 1945 Philadelphia birth cohort (see

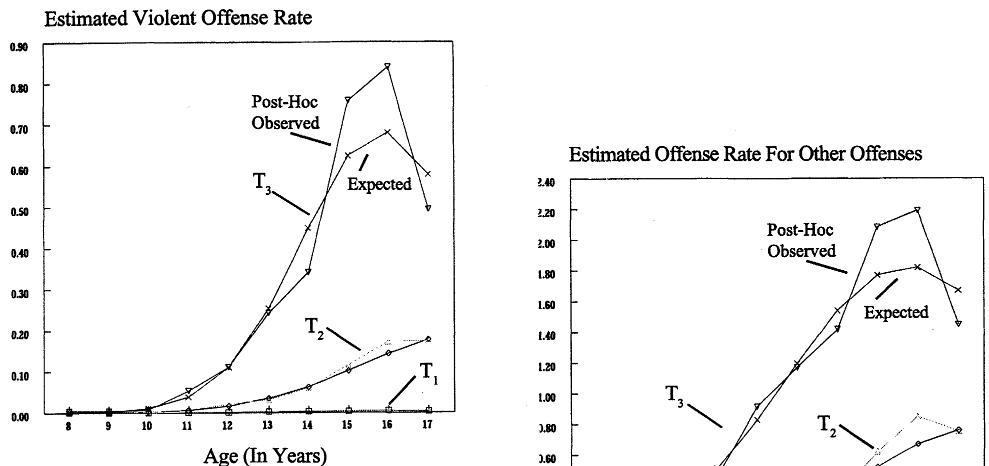




Estimated Mixing Weights

Component	π
T_1	.813
T_2	.157
T_3	.030

Figure 3: Mixture of Violent and Other Offense Trajectories in 1945 Philadelphia Birth Cohort



Estimated Mixing Weights

Component	π
T_1	.832
T_2	.147
T_3	.021

Figure 4: Mixture of Violent and Other Offense Trajectories in 1958 Philadelphia Birth Cohort

Figure 3), approximately 81 percent of the individuals occupy the first trajectory, which has very low offense frequencies in both offending categories throughout the follow-up period. About 16 percent of the cohort, however, exhibit moderate levels of both violent and other kinds of offending activity. An extreme group, which comprises about 3 percent of the cohort, displays relatively high levels of both violent and other forms of criminal activity throughout the follow-up period. The results associated with the 1958 cohort (see Figure 4) lead to similar conclusions.

Based on these results, we infer that the offense trajectories in the two Philadelphia cohorts differ largely in level rather than in shape. Although there is clearly some shape variation in these trajectories, the dominant theme is one of level variation. Another important theme emerging from these two figures is that individuals who exhibit relatively low levels of violent activity also exhibit relatively low levels of other kinds of criminal activity. Conversely, individuals who exhibit relatively high levels of violent activity also exhibit relatively high levels of other kinds of criminal activity. These results are consistent with the predictions of theorists who anticipate substantial generality in criminal behavior (Gottfredson and Hirschi 1990; Donovan and Jessor 1985; Rowe and Flannery 1994).

5. CONCLUSION

Social scientists have long been interested in the question of whether it is realistic to develop truly general theories of antisocial and criminal behavior. Some, such as Gottfredson and Hirschi (1990), claim that the bulk of the evidence suggests that individual rankings in behavioral tendencies are relatively time stable and that relative standing in the likelihood of committing one kind of crime is inextricably linked to relative standing in the likelihood of committing other kinds of crime. Others counter that proclivities for antisocial behavior are not necessarily so time stable (Loeber and LeBlanc 1990; Sampson and Laub 1993) and that it is misleading to try to theorize in such broad terms (Gibbons 1994; Gibbs 1987).

Like other investigators (see, e.g., Blumstein, Cohen, and Farrington 1988; Moffitt 1993; Farrington 1986), we think that studies that simultaneously (1) examine the development of antisocial activity over the life span and (2) investigate how different kinds of antisocial behaviors are linked to each other at different stages of the life span are likely to be valuable in these ongoing debates. In this article, we have attempted to highlight methods and produce descriptive evidence that will promote discussion and exploration of these issues. While our analysis of two general population data sets does not fully answer questions like those posed above, the substantive conclusions are provocative. Still, many more analyses like the one conducted here are necessary before the broader questions driving debates in this area can be resolved.

Two major themes emerged from our analysis. First, there is strong evidence that variation in criminal activity over the juvenile years is due to relatively stable individual differences in the tendency to offend rather than to fundamental heterogeneity in the shape of age-crime curves across the population. We base this conclusion on our finding that individuals who exhibit high levels of criminal activity at one point in time tend to exhibit high levels of criminal activity at other points in time as well. We wish to emphasize that our study is not the first to generate this result. Rather, our analysis simply replicates the finding of Nagin and Land (1993), who found evidence of pronounced age-crime curves of similar form across the population they studied. Second, the analysis presented here suggests that individuals who tend to exhibit relatively high (low) levels of violent criminal activity also tend to exhibit relatively high (low) levels of other forms of criminal activity.

We also believe that this analysis illustrates the usefulness of latent class and finite mixture models for the study of how behaviors develop over time. The approach taken here, based on the earlier efforts of Nagin and his colleagues (see, e.g., Nagin and Land 1993; Land et al. 1996; Roeder et al. 1999; Nagin 1999), allows for the identification of subgroups of individuals who display similar patterns of behavior on two outcomes over time. Application of this approach to other disciplines and questions could refine our thinking about the identification of individuals at risk for adverse life outcomes. For example, in the area of mental health, researchers could use methods based on this framework to study the joint temporal distribution of drug use and personality problems. In general, we suspect that questions about the relationship between different problem behaviors at different points in the life span can be productively investigated with the methods discussed here.

As suggested above, there are many substantive questions left unresolved by this work. For example, what would happen if we studied self-reported criminal activity instead of official record-based measures? Would we have obtained similar results if we had categorized offending behaviors differently? What would we find if we took time off the street (due to incarceration, death, etc.) into account? Is it possible to prospectively identify who is likely to embark on a given trajectory with any accuracy? Would we obtain similar results with other general population and criminal offender data sets? Would we obtain similar results for data sets that include females? We think all of these questions (and many others that we have not listed) are appropriate targets for future study.

NOTE

1. A number of other formulas for calculating this have been suggested in the literature. The formula here is from Wasserman (1997).

REFERENCES

Blumstein, A., J. Cohen, and D. P. Farrington. 1988. "Criminal Career Research: Its Value for Criminology." *Criminology* 26:1-35.

Blumstein, A., J. Cohen, J. A. Roth, and C. A. Visher, eds. 1986. *Criminal Careers and "Career Criminals."* Vol. 1. Washington, D.C.: National Academy Press.

Clogg, C. C. 1995. "Latent Class Models." 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.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood Estimation From Incomplete Data via the EM Algorithm (With Discussion)." *Journal of the Royal Statistical Society, Series B* 39:1-38.

Donovan, J. E. and R. Jessor. 1985. "Structure of Problem Behavior in Adolescence and Young Adulthood." *Journal of Consulting and Clinical Psychology* 53:890-904.

Donovan, J. E., R. Jessor, and F. M. Costa. 1988. "Syndrome of Problem Behavior in Adolescence: A Replication." *Journal of Consulting and Clinical Psychology* 56:762-65.

Elliott, D. S., D. Huizinga, and S. Menard. 1989. *Multiple Problem Youth*. New York: Springer-Verlag.

Everitt, B. S. and D. J. Hand. 1981. *Finite Mixture Distributions*. London: Chapman and Hall.

Farrington, D. P. 1986. "Age and Crime." In *Crime and Justice: A Review of Research*, vol. 7, edited by M. Tonry and N. Morris. Chicago: University of Chicago Press.

Farrington, D. P., R. Loeber, D. S. Elliott, J. D. Hawkins, D. B. Kandel, M. W. Klein, J. McCord, D. C. Rowe, and R. E. Tremblay. 1990. "Advancing Knowledge About the Onset of Delinquency and Crime." In *Advances in Clinical and Child Psychology*, vol. 13, edited by B. Lahey and A. Kazdin. New York: Plenum.

Gibbons, D. C. 1994. *Talking About Crime and Criminals: Problems and Issues in Theory Development in Criminology*. Englewood Cliffs, NJ: Prentice Hall.

Gibbs, J. P. 1987. "The State of Criminological Theory." *Criminology* 25:821-40.

Gilmore, M. R., J. D. Hawkins, R. F. Catalano, L. E. Day, M. Moore, and R. Abbott. 1991. "Structure of Problem Behaviors in Preadolescence." *Journal of Consulting and Clinical Psychology* 59:499-506.

Gottfredson, M. R. and T. Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.

Hirschi, T. and M. R. Gottfredson. 1983. "Age and the Explanation of Crime." *American Journal of Sociology* 89:552-84.

Huesmann, L. R., L. D. Eron, M. M. Lefkowitz, and L. O. Walder. 1984. "Stability of Aggression Over Time and Generations." *Developmental Psychology* 20:1120-34.

Jessor, R. 1992. "Risk Behavior in Adolescence: A Psychosocial Framework for Understanding and Action." *Developmental Review* 12:374-90.

Johnson, N., S. Kotz, and A. W. Kemp. 1992. *Univariate Discrete Distributions*. New York: John Wiley.

Kass, R. E. and A. E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90:773-95.

Kass, R. E. and L. Wasserman. 1995. "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion." *Journal of the American Statistical Association* 90:928-34.

Land, K. C., P. McCall, and D. S. Nagin. 1996. "A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models With Empirical Applications to Criminal Careers Data." *Sociological Methods and Research* 24:387-440.

Land, K. C. and D. S. Nagin. 1996. "Micro-Models of Criminal Careers: A Synthesis of the Criminal Careers and Life-Course Approaches via Semiparametric Mixed Poisson Models With Empirical Applications." *Journal of Quantitative Criminology* 12:163-91.

Laub, J. H., D. S. Nagin, and R. J. Sampson. 1998. "Good Marriages and Trajectories of Change in Criminal Offending." *American Sociological Review* 63:225-38.

Loeber, R. 1982. "The Stability of Antisocial and Delinquent Child Behavior: A Review." *Child Development* 53:1431-46.

Loeber, R. and M. LeBlanc. 1990. "Toward a Developmental Criminology." In *Crime and Justice: A Review of Research*, vol. 11, edited by M. Tonry and N. Morris. Chicago: University of Chicago Press.

Moffitt, T. E. 1993. "Adolescence-Limited and Life-Course Persistent Antisocial Behavior: A Developmental Taxonomy." *Psychological Review* 100:674-701.

Nagin, D. S. 1999. "Analyzing Developmental Trajectories: A Semi-Parametric, Group-Based Approach." *Psychological Methods* 4:139-57.

Nagin, D. S. and K. C. Land. 1993. "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed-Poisson Model." *Criminology* 31:327-62.

Nagin, D. S. and R. E. Tremblay. 1999. "Trajectories of Boys' Physical Aggression, Opposition, and Hyperactivity on the Path to Physically Violent and Non-Violent Juvenile Delinquency." *Child Development* 70:1181-96.

Osgood, D. W., L. D. Johnston, P. M. O'Malley, and J. G. Bachman. 1988. "The Generality of Deviance in Late Adolescence and Early Adulthood." *American Sociological Review* 53: 81-93.

Patterson, G. R. and K. Yoerger. 1993. "Developmental Models for Delinquent Behavior." In *Mental Disorder and Crime*, edited by S. Hodgins. Newbury Park, CA: Sage.

Piquero, A. R., A. Blumstein, R. Brame, R. Haapanen, E. P. Mulvey, and D. S. Nagin. 2001. "Assessing the Impact of Exposure Time and Incapacitation on Longitudinal Trajectories of Criminal Offending." *Journal of Adolescent Research* 16:54-74.

Roeder, K., K. Lynch, and D. Nagin. 1999. "Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology." *Journal of the American Statistical Association* 94:766-76.

Rowe, D. C. and D. J. Flannery. 1994. "An Examination of Environmental and Trait Influences on Adolescent Delinquency." *Journal of Research in Crime and Delinquency* 31:374-89.

Sampson, R. J. and J. H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6:461-64.

Tanner, M. A. 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 3d ed. New York: Springer-Verlag.

Tracy, P. E. and K. Kempf-Leonard. 1996. *Continuity and Discontinuity in Criminal Careers*. New York: Plenum.

Tracy, P. E., M. E. Wolfgang, and R. M. Figlio. 1990. *Delinquency in Two Birth Cohorts*. New York: Plenum.

Wasserman, L. 1997. *Bayesian Model Selection and Model Averaging*. Technical Report. Pittsburgh: Carnegie Mellon University, Department of Statistics.

Wolfgang, M. E., R. M. Figlio, and T. Sellin. 1972. *Delinquency in a Birth Cohort*. Chicago: University of Chicago Press.

Robert Brame is an assistant professor in the Department of Criminology and Criminal Justice at the University of Maryland, College Park. His work on this project was supported by a postdoctoral fellowship from the National Consortium on Violence Research. His research includes the study of legal issues, domestic violence, and juvenile delinquency. Some of his recent publications have appeared in Criminology, Journal of Quantitative Criminology, *and* Law and Society Review.

Edward P. Mulvey is professor of psychiatry and director of the Law and Psychiatry Program at Western Psychiatric Institute and Clinic at the University of Pittsburgh School of Medicine. His research focuses primarily on determining how professionals make judgments regarding the type of risk posed by adult mental patients and juvenile offenders and how to improve decision making about what interventions might be appropriate for these types of cases. Some of his recent publications have appeared in Archives of General Psychiatry, Social Psychiatry and Psychiatric Epidemiology, Law and Human Behavior, *and the* Virginia Journal of Social Policy and the Law.

Alex R. Piquero is associate professor in the College of Criminal Justice at Northeastern University. His research interests include criminological theory, crime over the life course, and quantitative research methods. His research has appeared in various publications, including Criminology, Journal of Quantitative Criminology, Journal of Research in Crime and Delinquency, Justice Quarterly, *and the* Journal of Adolescent Research.