SELECTIVE INCAPACITATION AND THE PHILADELPHIA COHORT DATA

by

Arnold Barnett and Anthony J. LoFaso

M.I.T. Working Paper #154484 March 1984

SELECTIVE INCAPACITATION AND THE PHILADELPHIA COHORT DATA

The University of Pennsylvania study of delinquency in a Philadelphia birth cohort has been described by Newsweek as "perhaps the most influential piece of criminal justice research in the last decade." (3/23/81) Many have construed the findings as showing that, if imprisonment were focused on the minority of offenders with especially bad "prognoses," the rate of crime could be reduced substantially. But others have taken the opposite view that the cohort data, far from endorsing such a "selective incapacitation" strategy, might actually provide strong evidence that such an approach is futile. Through some further analyses of the Philadelphia data, we attempt to clarify their policy implications.

ARNOLD BARNETT* and ANTHONY J. LOFASO**

*Sloan School of Management
Operations Research Center
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

**School of Business
Hofstra University
Hempstead, New York

ACKNOWLEDGEMENTS

To be supplied: the list grows daily.

INTRODUCTION

The more-than-doubling of the U.S. prison population in the last decade--from 196,000 in 1973 to 432,000 in 1983--has led to severe and often dangerous overcrowding. Yet there is little enthusiasm for dealing with the problem through massive prison construction; such activity takes considerable time, is enormously expensive, and evokes bitter controversy about where new facilities should be located. Against such a backdrop, any proposal that might allow the more effective use of existing prison space can count on a sympathetic hearing.

Among such proposals, perhaps the most widely discussed is the call for "selective incapacitation." The advocates of this position contend that some offenders commit crimes at far higher rates than others, and that it is possible to characterize that subset of offenders whose prognoses are the bleakest. Moreover, the argument goes, the prisons can do little either to deter people from criminal careers or to rehabilitate those already engaged in them. (I.e. the only major mechanism by which imprisonment reduces crime is through interrupting the activities of those behind bars). Given these viewpoints, the proponents of selective incapacitation would focus scarce prison resources on a minority of offenders: those deemed likely to amass the worst future records.

As of now, there is no concensus that selective incapacitation is a viable strategy. Some people believe it is inherently unjust to base punishment not on what an offender has done, but rather on what he is expected to do (e.g. Von Hirsch (1976)). Many are troubled because, given the imperfection of any predictive technique, some people would inevitably be punished in anticipation of crimes they would never actually have committed. Perhaps most important, disagreement persists on the fundamental question of whether it is really possible, using accessible and appropriate information, to make even a moderately-accurate forecast of an offender's future behavior.

This last question is essentially an empirical one. Much of the initial and continuing support for selective incapacitation arose from the University of Pennsylvania study of delinquency in a Philadelphia birth cohort (Wolfgang, Sellin, Figlio (1972)). The analysis revealed that 6% of the cohort's members accounted for more than half the entire group's police contacts for juvenile delinquency. And this 6% was a small minority even within the offender population. Many people saw in such statistics a strong argument for "cracking down" on the "chronic" offenders; so many, that, as noted in the abstract, Newsweek viewed the Philadelphia cohort study as "perhaps the most influential piece of criminal justice research in the last decade." (3/23/81).

But the heterogeneity of the criminal class implied by these Philadelphia numbers, although a necessary condition for a selective incapacitation strategy, is not sufficient to ensure its success. To exploit such diversity, one must have some reliable means of predicting in advance who the worst offenders are likely to be. Wolfgang et al did not identify any such procedure and Blumstein and Moitra (1980), having examined the Philadelphia data, suggested it might not be possible to do so.

To put it briefly, it is conceivable that the Philadelphia cohort data, so influential in increasing the credibility of selective incapacitation, might actually contain strong evidence that the approach is futile. At the same time, nothing presented so far precludes an opposite possibility: the data may offer far deeper support for the theory is now realized. To "get to the bottom" of what the Philadelphia data are saying would require a more detailed analysis of certain aspects of the records than has yet been undertaken. We attempt in this paper to perform such an analysis.

We will be describing a series of investigations linked by the theme: to what extend does a youth's police record foreshadow his future involvement in delinquency? The studies will be based on the Philadelphia cohort data, generously provided by colleagues at the University of Pennsylvania. The aim is not to find some "optimal" predictive formula, but rather to get some sense of how powerful an attainable formula might be.

To be sure, the most recent research on selective incapacitation has taken place outside of Philadelphia, perhaps most notably at the Rand Corporation. (e.g. Chaiken and Chaiken (1982); Greenwood (1982)). Such work is important, and deserves the prominent place that it occupies in the ongoing policy debate. But it is not to denigrate these efforts to point out (as we will) that, in some respects, the Philadelphia data might be uniquely valuable. Thus, while a Philadelphia study would by no means supplant those conducted elsewhere, it need not be eclipsed by them either.

The findings that will unfold in forthcoming sections have a bittersweet quality. We will suggest that, from relatively simple statistics about past juvenile crimes, one can make discriminating statements about the pattern of future ones. But we will also suggest there might be somewhat less diversity among Philadelphia delinquents than certain summary statistics may have led us to believe. Thus, while it might be feasible to identify high-rate offenders, they might be sufficiently similar to other candidates for imprisonment that the value of doing so is limited.

We begin our detailed discussion in the next section, where we review the Blumstein-Moitra analysis of the Philadelphia cohort data. Then in Section III, we place the prediction problem in a more general framework. Some exploratory analyses in Sections IV-IX culminate in a more formal modeling exercise in Section X. The remainder of the main text considers the question "Why Philadelphia?" and discusses some strengths and limitations of this work. Various technical details appear in Appendices A-C.

II. THE BLUMSTEIN-MOITRA PAPER

The original Philadelphia cohort study concerned 9945 males born in 1945 who lived in the city between ages 10 and 18. It emerged that 35% of the youths were arrested* at least once by their 18th birthdays, and that the most active 6%, who averaged eight arrests apiece, sustained slightly over half of the cohort's police bookings. Wolfgang et al defined these 6%, all of whom had at least five arrests, as "chronic offenders"; testifying before a Congressional committee, Professor Wolfgang stated that their activities "loudly claim attention for a social action policy of intervention."

But, in an intriguing paper, Blumstein and Moitra (1980) wondered just what public actions would successfully restrain the chronic offenders. They presented an analysis that suggested the length of a youth's police record, the most obvious indicator of his propensity for future delinquency, was virtually uncorrelated with this number of <u>further</u> juvenile arrests. (The only major exceptions to this pattern were first-offenders, almost half of whom were never rearrested.)

---

* We follow Blumstein & Moitra in using the word "arrests"; Wolfgang et al preferred the phrase "police contacts."

More specifically, Blumstein and Moitra found the cohort data consistent with a simple probabilistic model. Given that a youth has just been arrested a k^{th} time ($k \geq 3$), there is roughly a 72% chance he will go on to a k+1st juvenile arrest (and thus a 28% chance that he will not.) If he is arrested a k+1st time, there is again a 72% chance that he will be a k+2nd, etc. Under this geometric progression, his total number of future juvenile arrests has an expected value of $.72/(1-.72) = 2.57$, <u>for every value of k above two</u>. Even for k=2, the corresponding number is only slightly lower at 2.32.

The upshot is that, given a set of delinquents with two or more arrests, predicting which of them will lengthen their records the furthest is rather like predicting which rolls of the dice will come up seven. Discussing the policy implication of their finding, Blumstein and Moitra stated that "if we imprison all people who have already been arrested three times, we avert 2.57 future arrests per prisoner. But if we restrict imprisonment to the more chronic offenders who have already been arrested seven times," then "we will (also) avert only 2.57 future arrests per prisoner." Under such a pattern, "locking up the 'career criminal' averts no more crimes than locking up any other persister."

The Blumstein-Moitra calculation makes the exceedingly important point that, at the aggregate level at which the Philadelphia data are generally discussed, they are consistent with an interpretation wholly uncongenial to selective incapacitation. But this is <u>not</u> to say that it shows that the theory cannot be implemented. Even if k itself lacks predictive value, the <u>rate</u> at which the k arrests were accumulated might have some. Perhaps <u>changes</u> in the headways between consecutive arrests convey some information, as might the delinquent's age.

Indeed, the Blumstein-Moitra finding does not even foreclose the possibility that k actually is a useful guide to future behavior. It is important to bear in mind that the statistic 2.57 pertains not to a youth's total number of future crimes but only to those while he is still a juvenile (i.e. through his 18th birthday). This proviso could be highly important, as a simple example suggests.

Suppose that among youths with at least seven policy bookings, the average age at the time of the seventh is 17. Then, if possible detention is neglected, the Blumstein-Moitra statistic implies that, for the remainder of their days as juveniles, these boys will commit known offenses at an average annual rate of roughly 2.57. Now suppose that the average age of a youth at the time of his third police contact is only 15. If these boys also will average 2.57 future juvenile arrests, the rate of such arrests per year would be something like 2.57/3 = .86. In this hypothetical situation, imprisoning people for a year after their seventh arrest would have about three times the incapacitative effect of doing so for those from whom k=3.

In summary, Blumstein and Moitra considered only one possible predictor of further delinquency. And they focused on predicting "future juvenile crimes" rather than "future juvenile crimes per year," while it is the latter quantity that enters the cost-benefit analysis tied to selective incapacitation. Thus it is probably prudent to view the Blumstein-Moitra result more as a conjecture worthy of further investigation than as a definitive conclusion in its own right. (Blumstein and Moitra themselves suggested their work was meant more as a catalyst than anything else.)

Starting in the next section, we present the results of some further analyses.

III. FRAMEWORK OF THE STUDY

Traditionally, authors describe their method of analysis before discussing its possible weaknesses. But one of our simplifying assumptions might seem so unnerving that it is worth examining at the outset. Throughout this paper, we will speak generically of crime and not consider directly the varying seriousness of different offenses. Such a convention makes the calculations easier to perform and interpret, but it has the unpleasant consequence of obliterating the distinction between a petty larceny and a rape.

The simplification we are making, far from being unprecedented in criminal justice research, implicitly is present in both the use of the famous 6% statistic and in the Blumstein-Moitra analysis. Indeed, to the extent that we are trying to resolve an existing controversy, there is no need to disaggregate the data by crime-type. But neither the authors nor, we presume, many readers will find this rationale wholly satisfactory.

The real fear is that a negative correlation may exist between the length of a youth's juvenile police record and the seriousness of his crimes. Given that detention exists, one might speculate that only boys whose crimes are trivial are free long enough to accumulate long records. There is proper concern about any analysis under which eight loitering offenses could seem worse than six armed robberies.

Fortunately, the Philadelphia data -- which record arrests both by date and offense type -- allow some empirical explorations. A crude dichotomy might classify crimes as either "serious" or "minor": the former might consist of the F.B.I. Index offenses plus arson, kidnapping and malicious mischief, while the latter would comprise all others. Table 1 uses such a partitioning rule to relate total record length to crime seriousness for the Philadelphia Delinquents.

TABLE 1: RECORD LENGTH VS. CRIME SERIOUSNESS FOR ACTIVE PHILADELPHIA DELINQUENTS

TOTAL NUMBER OF OFFENSES	PERCENTAGE THAT WERE SERIOUS
3-4	48.4
5-6	51.4
7-8	52.6
9-10	55.2
11-12	49.9
13 OR MORE	55.0

NOTE: See Appendix A for details of the calculations.

The table depicts a striking pattern: the virtual absence of any relationship between the number of arrests and percentage in the worse cateogry. (Like Blumstein and Moitra, we will focus on youths with three or more arrests.) And in Appendix A, we indicate that this outcome would not change much if we raise the threshold that separates "serious" and "minor" crimes.

In the aggregate, therefore, a rough proportionality seems to exist between a delinquent's record length and the total known harm he has caused to society. Under the circumstances, the consequences of not considering crime-type, though surely not zero, are probably less than we might initially have expected. Hence our simplifying assumption has more than just precedent working in its behalf.

Now we begin the narrative description of our analytic approach. If crime-type is neglected, a youth's police record at the time of his k^{th} apprehension would consist of:

(i) k itself

(ii) present age, a_k

(iii) ages at all previous arrests, $a_1, a_2, \ldots, a_{k-1}$

The 1945 cohort data include the race of each boy and some sociometric statistics concerning 10% of the group. But using some such factors is clearly impermissible in sentencing decisions, and the status of the others is questionable. Thus our focus will be on the predictive power of the offense record <u>in itself</u>.

The general goal is to consider predictive equations of the form:

$$\dot{s} = f(a_k, k, \dot{k}, \ddot{k}) + \varepsilon \tag{1}$$

where

$\dot{s}$ = future juvenile arrest <u>rate</u> over some specified period

$\dot{k}$ = a measure of average arrest rate up through the k^{th} booking

$\ddot{k}$ = a measure of trends in the intervals between arrests (i.e. something akin to the "second derivative" of k with respect to age)

ε = a zero-mean "random error" term

We are disinclined to use predictor variables that have nothing to do with behavior near a_k (e.g. a_1). Hence the absence of such variables in (1).

Thus formulated, the forecasting problem seems amenable to some multivariate statistical technique, probably multiple regression analysis. But is essential that we consider this possibility with caution. Any regression analysis depends on a series of strong assumptions, the imperfections of which could reverse the signs and distort the magnitudes of the key coefficients. This is not merely a hypothetical danger; one of the authors, for example, has suggested such reversals might well have occurred in certain regressions about job discrimination and about the deterrent effect of capital punishment (Barnett (1981), (1982)).

Inspecting (1), one can easily see some potential problems in calibrating it. Even is $\dot{s}$ is related to k, no real theoretical basis exists for assuming the dependence is linear or, for that manner, of any other particular form. The same is true of a_k, for which the relationship need not even be monotonic (e.g. all other factors held constant, $\dot{s}$ might increase with a_k at first and then decrease as a_k grows further).

Furthermore, the variables k and $\dot{k}$ are clearly correlated: the record is truncated at age 18, and to have numerous arrests in a limited time one must experience them at a rapid rate. Such correlations are hardly unknown in regression analysis, but they can yield unstable parameter estimates in the best of circumstances. And if the regression model has other flaws, its attempt to "partition" an effect among correlated variables could fail.

The assumptions about the random-error term, ε, are highly important, for they have strong implications about the accuracy assigned to individual forecasts. The ε's reflect the stochastic processes under which active offenders sustain arrests and by which youths "drop out" of the delinquent population. It is not a priori clear which probability models best describe these processes; indeed, to rush to a specification of the ε's would be to dismiss at the outset some of the subtlest problems in predicting future delinquency.

Let us be clear: we have no antipathy to using familiar multivariate methods. But we believe strongly that, prior to doing that, one should explore the data in considerable detail. Only then is one in a position to use the methods in ways that yield results that are genuinely trustworthy.

In the next section, we start our exploration of the cohort data with a modification of the Blumstein-Moitra calculation. The subsequent five sections consider -- first singly and then in combination -- the predictive value of k, $\dot{k}$, $\ddot{k}$, a_k, and the various influences on ε. Though the sequence of tests and calculations might seem increasingly tedious, it is essential to justifying the simple model of Section X. Please bear with us.

IV. RECORD LENGTH AS A FIRST PREDICTOR

Although Blumstein and Moitra showed that the prior number of juvenile arrests is unrelated to the future number, record length could still be associated with the future arrest rate. We consider that possibility now. We will refer to those delinquents with at least k arrests as the "k-subgroup," and to their police encounters after the k^{th} and before age 18 as "subsequent arrests."

γ_k, a useful first statistic about the k-subgroup, is given by:

$$\gamma_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \frac{n_{ik}}{18-a_{ik}} \tag{2}$$

where N_k = number of members of the k-subgroup

n_{ik} = number of subsequent arrests by the i^{th} member

a_{ik} = age of the i^{th} member when he entered the k-subgroup

Note that, if possible detention is neglected, $n_{ik}/(18-a_{ik})$ is simply member i's rate of subsequent arrests. γ_k is the arithmetical average of the subsequent arrest rates of all k-subgroup members. Thus defined, γ_k has the following interpretation: if a randomly-chosen member of the k-subgroup is locked up from his k^{th} arrest through age 18, γ_k is the expected number of arrests averted per year.

The remark about neglecting possible detention, though useful for the flow of the last paragraph, would be pernicious if the basis of the calculations. Of those youths with three or more arrests, roughly 30% spent at least some time in a correctional institute. It appears, moreover, that average detention among those who experienced it was about 1.5 years. Ignoring such confinement would give a substantially biased view of how its recipients behaved while free.

For those delinquents with detention time, we calculated subsequent arrest rates under procedures described in Appendix B. Combining such adjusted statistics with the rates computed under (2) for the youths not detained, we obtained the numbers that appear in Table 2.

TABLE 2: FUTURE ARREST RATES VS. PRESENT RECORD LENGTHS AMONG PHILADELPHIA DELINQUENTS

k	N_k	γ_k
3	1033	1.15
4	708	1.43
5	534	1.52
6	386	1.79
7	303	1.92
8	224	2.08
9	175	2.25

where k = number of arrests thus far

N_k = number of youths with at least k arrests

γ_k = average subsequent arrest rate in "k-subgroup"

Table 2 displays a steady (and, curiously, a roughly linear) pattern under which γ_k grows with k. There are no exceptions to the upward trend. It seems clear that, once future arrests are normalized by time of exposure, the number of bookings thus far <u>does</u> have some predictive value.

However, when k goes from three to nine, γ_k only rises by a factor of two. One might sense that this means k is only a weak predictor, explaining but a small fraction of the variance in future arrest rates. Such suspicions are fueled by the well-known summary statistic: if 6% of the cohort absorbs half its arrests (i.e. as much as the other 94%), then something like a factor of <u>sixteen</u> separates the "chronic offenders" from the other Philadelphia youths.

But there is an important distinction to be made, related to which population of youths is the proper focus of interest. Prison terms are not meted out to those never arrested or, with rare exceptions, to first or second offenders. In considering youths with at least three arrests, Blumstein and Moitra were thus restricting attention to a key subset of juveniles: those for whom punishment decisions are relevant.

The data show that only about 12% of the Philadelphia cohort members were arrested at least three times. Thus, among the serius candidates for detention, (hereafter the "active offenders") the "chronic offenders" of Wolfgang et al were really just the busier half.*

When, as in Table 2, those with fewer than three arrests are excluded, the familiar 6%|52% disparity takes a less dramatic form: the top 52% of the active offenders sustained 70% of the whole group's arrests. Correspondingly, the original factor of 16 by which "chronic offenders" differed from others is reduced to a factor of two. Given this last ratio, the predictive power of k implied by Table 2 is not nearly as unimpressive as it might have first seemed.

We should not jump to conclusions about the prognostic value of k. The apparent importance of record length could well be spurious, caused solely by its correlation with a genuine predictor variable. This possibility is among those considered in subsequent sections.

---

* As explained in Appendix A, the N_k's in Table 2 exclude curfew offenses; hence they imply slightly lower percentages than the traditional figures being cited here.

But the rudimentary calculations performed so far have made a significant point. The pessimism engendered by the Blumstein-Moitra findings might have been premature; even in the Philadelphia data the authors studied, the future rate of arrests is systematically related to the existing number. Hence the outcome least favorable to selective incapacitation -- the utter uselessness of the present offense record in projecting the future one -- can already be excluded.

V. PAST ARREST RATE AS A PREDICTOR

It is natural to speculate that one's rate of arrests in the past bears a close relationship to that in the future. We could explore this possibility with an analogue to Table 2, computing average future arrest rates over groups of offenders whose past arrest rates differ. But proceeding in that way would raise a dilemma. We noted earlier that k and $\dot{k}$ are correlated; thus, even if each seemed to have predictive value, it would be unclear how to combine the two findings. The real issue is how much incremental power one gets using $\dot{k}$ and k together rather than just one of them.

To "disentangle" such correlated effects, it is useful to hold one of the factors constant and then assess the importance of variations in the other. In the present context, we might choose a fixed value of k (e.g. five) and consider whether, among those who just entered the k-subgroup, the rate at which they "earned" membership suggests their future rate of arrests. (Later, we will discuss the reverse calculation in which $\dot{k}$ is held constant and k is varied.)

The k-subgroup can be divided into the desisters, whose delinquent careers seem to end with the k^{th} booking, and the persisters, who go on to subsequent arrests. Two questions worth considering are:

(i) Is a delinquent's $\dot{k}$-value related to the probability he will desist?

(ii) Among those who persist, is $\dot{k}$ correlated with the rate at which they do so?

These two questions are quite distinct. If, as has been suggested, offenders commit crime at a constant rate until some unpredictable moment when they cease activity altogether, the answers to (i) and (ii) would be "no" and "yes" respectively.

As posed, these questions can be handled with nonparametric methods. To answer the first, one could rank the k-subgroup members according to their $\dot{k}$-values, and then use a Wilcoxon rank-sum test to see whether those who desisted were spread uniformly through the ranks. To deal with the second, one could rank the persisters on two criteria: their $\dot{k}$-values, and their subsequent arrest rates. A Spearman rank-correlation test could indicate whether the ranking on the second dimension is foreshadowed by that on the first.

Actually, this Wilcoxon test on desistance has flaw, related to a distinction between "true" and "false" desistance. While the problem is correctible, it pertains to a major concern about selective incapacitation. Thus it is not merely digressing to turn to the problem at once.

VI. FALSE DESISTANCE

Even if a youth has a clean police record between his k^{th} arrest and age 18, it is not certain that he "went straight" at a_k. He might have committed later crimes for which he wasn't caught or gone through a temporary lull in activity. Indeed, as a_k gets closer to 18, attributing significance to the absence of further arrests becomes farfetched. (Who would argue that an offender had "reformed" because he had no juvenile bookings after 17.8?)

In other words, we must distinguish between the "true" desistance caused by retirement from crime and its "false" counterpart caused by terminating the record at the onset of adulthood. And we must try to estimate how observed desistance is divided between the two categories.

The following data concern those in the 1945 cohort arrested a fifth time. (We concentrate now--and later--on the fifth arrest because that is the one that places a youth in Wolfgang's "chronic offender" group.)

TABLE 3: DESISTANCE AMONG FIFTH ARRESTEES

Age at Fifth Arrest(a_5)		Percentage who "Desisted" Thereafter(a_5)	
14-14.5	(22)	18.4	(4)
14.5-5	(47)	8.5	(4)
15-15.5	(50)	12.0	(6)
15.5-16	(63)	17.4	(11)
16-16.5	(51)	19.6	(10)

NOTE: The numbers in parentheses show the actual number of "fifth offenders" in each category.

At 14.5%, the average desistance rate for the youths in Table 3 is only about half that for all fifth arrestees (28%). It follows that, among those arrested a fifth time <u>after</u> 16.5 (i.e. nearly all the chronic offenders not in the table), the rate of desistance is roughly 37%. And, within the table itself, the fraction without further juvenile bookings generally grows as the youths get older. (The only exception to this pattern is the youngest category, which has few members.) Overall, then, there seems a clear tendency for the observed desistance rate to <u>increase</u> with age.

One possible interpretation of the data is that, the younger one is at the fifth arrest, the more "incorrigible" one is. But it is also possible that the pattern is artifactual because, the older one is at the fifth arrest, the easier it is to desist falsely.

Suppose we consider the extreme hypothesis that all the desistance in the Table 3 is false. The underlying idea is that, even in the midst of an active career, there will be particular periods in which no arrests occur. Among those last arrested at a_5, was the age range (a_5, 18) just such a quiescent period and nothing of genuine importance?

To try to answer that question, one must make some probabilistic assumptions about patterns of personal arrest. Many criminal justice researchers (e.g. Avi-Itzhak and Shinnar (1973), Blumstein and Nagin (1978), Chaiken and Rolph (1980), Greenwood (1982), and Maltz and Pollock (1980)) have assumed that, while free during active careers, offenders accumulate arrests under Poisson processes with constant means. (The mean itself can vary from offender to offender.) And, later in this paper, we will present evidence of consistency between the Poisson model and the Philadelphia data. We will therefore use a Poisson formulation in estimating the amount of false desistance.

Consistent with the "no true desistance" hypothesis, suppose λ is a youth's Poisson arrest rate both from the start of his career through a_5 and between a_5 and 18. Then Q, the probability of no arrests in (a_5, 18) would follow:

$$Q = e^{-\lambda(18-a_5)}$$

We will refer to Q as the probability of false desistance, because it treats the absence of arrests in (a_5, 18) as a fluctuation in an ongoing process rather than a sign of real change.

To test the null hypothesis that true desistance is nonexistent, one could estimate the Q's for the various "fifth offenders" and then, through summing the individual Q_i's, find the expected number of false desisters if the hypothesis were true. Under an independence assumption, the variance of that total number would be approximated by the sum of the $Q_i(1-Q_i)$'s. A relevant statistical test would see whether the actual number of desistors was sufficiently close to the projected number.

The statistical test we performed is described in Appendix C. It turns out that, if all desistance were false, one would have expected 27 of the 233 youths in Table 3 to have had no further juvenile bookings. The actual number, 35, exceeds the expected level to a statistically significant extent. But while we can reject the hypothesis that true desistance never occurs, the excess of 35 over 27 suggests that, of the 233 delinquents, only about <u>nine</u> really desist at the fifth arrest.* The original dropout rate of 28%, therefore, might overstate the true rate of desistance by something like a factor of <u>eight</u>.

Like the finding that γ_k grows with k, this outcome works against a potential objection to selective incapacitation. The more youths who abandon delinquency at random times (or, in Wolfgang's words, enter "spontaneous remission"), the more frequently will a selective incapacitation scheme err egregiously. The apparent rarity of true desistance suggests that, in the Philadelphia data at least, this particular danger might not be acute.**

---

* 27 false desisters out of 233 youths implies a false desistance rate of 11.6%. If there are x genuine desisters--and they are scattered through the population--the total number of youths with no further arrests could be approximated by $x + .116(233-x)$. Equating that quantity to 35 (the observed number of desisters) yields $x \approx 9$.

** The reader might be wondering if the same pattern would have arisen had we focused on those with (say) seven arrests so far. That question in turn raises a broader one: is it reasonable, in considering a given hypothesis, to examine only a subset of those Philadelphia data related to it? We believe the answer is "yes," provided the subset was chosen in advance under plausible, easily-stated rules. There is no reason to expect such a "sampling" procedure to yield misleading results, and the alternative would entail an unwieldy set of correlated findings.

VII. RETURN TO $\dot{k}$

Back on page 17, we first raised the issue of false desistance in connection with a specific question: are $\dot{k}$-values correlated with the chances of further arrest? Among youths of comparable ages, those whose $\dot{k}$'s are lower would seem more prone to desist falsely. Thus a negative correlation between $\dot{k}$ and the desistance rate might be expected for purely technical reasons; this circumstance could clearly distort the Wilcoxon test we had been contemplating.

Adjusting the Wilcoxon procedure to "weed out" the potential bias would be feasible. But given evidence that true desisters are few and far between, attempts to identify them in advance do not seem promising. Thus we will not pursue the matter further.

There is still, however, the question of whether, among persisters, the $\dot{k}$-value is related to the rate of future arrests. To answer it, we considered the 200 youths in Table 3 who did persist, and calculated the Spearman rank-correlation coefficient of their past and future arrest rates. The coefficient was only .178 but, given the large sample size, it was significantly greater than zero under a one-sided test at the 5% level.

Because this last calculation broke the time line at the fifth arrest (i.e. k=5 for all youths), the significant Spearman statistic implies that $\dot{k}$ has marginal predictive power on its own. But an interesting if disheartening pattern emerges if, as in the previous section, we segment the delinquents by their ages at fifth arrest.

TABLE 4: CORRELATION OF PAST AND FUTURE ARREST RATES, BY AGE AT THE FIFTH ARREST

AGE AT FIFTH ARREST		SPEARMAN RANK-CORRELATION COEFFICIENT
14-14.5	(18)	-.422
14.5-15	(43)	-.078
15.15.5	(45)	.126
15.5-16	(52)	.399
16-16.5	(42)	.484

Of the five correlations presented above, only the last two differ significantly from zero (and, indeed, do so at the 1/2 of 1% level). But there are no exceptions to the trend under which the coefficient gets larger with age.

We might have anticipated such a directional pattern on statistical grounds. By chance alone, some youths will reach their fifth arrests artificially early; their k-values would in consequence by misleadingly high. The proportion of youths of this type would presumably get <u>larger</u> as the age range gets lower. The "regression-to-the-mean" effect associated with studying the youths further would, therefore,have progressively stronger influence on the earlier Spearman coefficients.

But it would be going too far to assert that the pattern lacks behavioral significance. It is certainly conceivable that young delinquents are especially volatile and unpredictable, while older ones exhibit greater "stability." Thus a predictor variable of considerable power for the latter group could be wholly ineffective for the former.*

---

* In the last section, the rejection of the "no true desistance" hypothesis depended crucially on its failure in the age range (14, 14.5). Thus, beyond the Spearman calculation above, other evidence strengthens the suspicion that the youngest delinquents are the most unpredictable.

VIII. ANOTHER LOOK AT k

We have uncovered evidence that $\dot{k}$ has predictive power in its own right, at least among older delinquents. Now we consider whether k itself has genuine value. Or, to raise the issue more pointedly, did k attain its successes in Table 1 only because it was "hitchhiking" on the predictive power of its correlate $\dot{k}$?

Once again, we proceed with a nonparametric method. We contrast pairs of youths whose ages and $\dot{k}$-values are very close but who, because they have been active for different amounts of time, have total records of unequal lengths. We compare the $\dot{s}$-values for the two "matched" delinquents. If k has inherent power then we would expect that f, the fraction of pairings in which the boy with the higher k has the higher $\dot{s}$, would significantly exceed .5.

To implement the test, we focus on the narrow age-range (15.25, 15.75), and consider youths arrested during that period for at least the fourth time. (For those arrested fewer than four times, $\dot{k}$-estimates are not very trustworthy.) The age range selected has the twin advantages of avoiding the especially unpredictable "early chronic offenders" and of allowing a long enough follow-up period that the $\dot{s}$-values are not wracked with instability.

Within the relevant population, we ranked the youths by their $\dot{k}$-values. The two highest became the first pairing; the third and fourth highest, the second, etc. We were therefore examining matched data of the following type:

Age at Trigger Arrest* (A_k)	$\dot{k}$-value	k	$\dot{s}$
15.42	1.87	6	2.16
15.51	1.86	4	1.40

NOTE: Provided that the delinquent has at least three previous bookings, the "trigger" arrest is the first in the age-range (15.25, 15.75).

Altogether, there were 46 useful pairings in the test set (i.e. pairings that remained when boys with equal k-values were excluded). In exactly 18 of them, the same youth had the higher k and the higher $\dot{s}$. The McNemar Test confirms that which is apparent to the naked eye: k-values are about as useful as fair coins in suggesting which youth will have the higher future arrest rate.

Across the various matches, $\dot{s}$ *was* clearly correlated with $\dot{k}$. The key point is that k had little *further* to contribute to the estimation of $\dot{s}$.

Seeing this outcome, we naturally returned to Table 2, in which γ_k rose steadily with k. For several subgroups of youths just arrested a k^{th} time, we computed the average arrest rate up to that point (i.e. the average $\dot{k}$). A monotonic relationship emerged in which $\dot{k}$ increased with k: for k=5, $\dot{k}$=1.72; for k=7, $\dot{k}$=2.29; for k=9, $\dot{k}$=2.92. Thus k could well have been serving as a "stand in" for the real predictor, $\dot{k}$.

The present results, therefore, do not contradict those in Table 2; rather, they probably explain them. Nor are the results inconsistent with the standard Poisson model, under which the past rate of bookings and not the actual number should generate the estimate of the future arrest rate.

Indeed, this last circumstance might lead readers to wonder why k was the first variable that we studied. We point out that we were pursuing a Blumstein-Moitra conjecture based on k and that, while defining and calculating $\dot{k}$ is not straightforward, doing so for k obviously is.

IX. THE SECOND DERIVATIVE AS A PREDICTOR

While one would not expect it under a stable Poisson process, the second derivative of the level of arrests ($\ddot{k}$) might turn out to have some predictive value. Perhpas steady drops in the headways between arrests often portend still further drops, while increases signal a gradual retirement from criminal activity. An empirical check on such possibilities seems worthwhile.

Now that k appears to be redundant of $\dot{k}$, a McNemar test about $\ddot{k}$ suggests itself. For paired youths with similar ages and $\dot{k}$-values, one could search for an association between $\ddot{k}$ and $\dot{s}$. If, in a percentage of matchings significantly different from 50, the higher $\ddot{k}$ and the higher $\dot{s}$ belong to the same boy, some predictive power of the second derivative would be suggested.

The questions arises of how $\ddot{k}$ should be measured. Given that we will be comparing youths whose $\dot{k}$-values are very close, differences in their $\ddot{k}$'s are tied systematically to those in their $\dot{r}$'s, where $\dot{r}$ follows:

$$\dot{r} = C((k-2)(a_k - a_{k-1}) + (k-3)(a_{k-1} - a_{k-2}) + \ldots + 2(a_4 - a_3) + 1 \cdot (a_3 - a_2))$$

$$\text{and } C = \frac{2}{(k-1)(k-2)}$$

Thus defined, $\dot{r}$ is a weighted average of the headways between arrests, with the emphasis steadily dropping as one moves back in the career. (The value assigned to C--one divided by the sum of integers from 1 to k-2--assures that the weighting factors add up to one.) For a fixed mean headway (i.e. a fixed $\dot{k}$), a higher $\dot{r}$ implies a <u>lower</u> $\ddot{k}$. In our matched comparisons, therefore, we can use $1/\dot{r}$ as a good surrogate for the second derivative.

Once again, we study those boys arrested in (15.25, 15.75) for at least the fourth time. In the 67 pairings we examined, the youth with the higher $\ddot{k}$ had the higher $\dot{s}$ in 37. 37/67 is a bit higher than .5, but the excess is nowhere close to statistical significance.

Given the high "noise level" surrounding this statistical test, some predictive power on the part of $\ddot{k}$ cannot be excluded by its results. But there are reasons to doubt such power could be substantial. Because relatively few crimes lead to arrests, headways between a youth's bookings could be far more responsive to his luck in the "arrest lottery" than to any changes in his crime commission rate. Thus trends in arrest intervals are necessarily of questionable importance, especially at the level of the individual offender.

In summary, we see no compelling reason for treating $\ddot{k}$ as a useful predictor.

X. A PREDICTIVE EQUATION

Recall that, early in the paper, we spoke of developing a forecasting model of the general form:

$$\dot{s} = f(a_k, k, \dot{k}, \ddot{k}) + \varepsilon$$

Now we can embark upon that task with some greater confidence about how to proceed.

Although hardly definitive, the findings in Sections IV - IX suggest that:

(i) In predicting $\dot{s}$ based on k, $\dot{k}$, and $\ddot{k}$, only the second of these three variables is of fundamental importance.

(ii) The true rate at which youths "drop out" of delinquent careers, is probably rather low. Thus, if the arrest rates of active offenders are stable over time, an estimate of $\dot{s}$ need not depend critically on the length of the period of prediction.*

---

* Of course, in restricting our attention to juvenile records, we are considering time frames of at most a few years.

(iii) Young delinquents may be more unpredictable than their "elders." It is therefore possible that, below some threshold value of a_k, any forecasting equation could prove ineffective.

Consistent with these hypotheses is a simple formula that relates $\dot{s}$ to $\dot{k}$:

$$\dot{s} = \alpha + \beta(\dot{k}-\mu) + \varepsilon \qquad \text{where } C < a_k < D \tag{3}$$

where μ = the average $\dot{k}$-value in the population under study

ε = a zero-mean "random noise" term

C and D are age limits specified in advance

α and β are constants to be calibrated from data

The linearity assumption in (3) does not seem especially controversial; any other functional form would seem harder to defend. (Note that the models $\dot{s}=\dot{k}$ and "$\dot{s}$ unrelated to $\dot{k}$" are special cases of (3).) More noteworthy is the relationship of (3) to the phenomenon of "regression to the mean."

Even if delinquents sustain arrests under stable Poisson processes (as we will assume here), the time until a youth's k^{th} arrest is subject to considerable randomness. Thus, among the $\dot{k}$-values for a given set of offenders, a good fraction generally exceed the "true" arrest rates (the λ's), while another large fraction fall below them. If the youths are ranked according to their $\dot{k}$'s, those whose arrest rates have been overestimated gravitate upward in the rankings; those in the opposite situation fall towards the bottom.

It follows that, if the boys are studied for a subsequent period, those with the highest (lowest) $\dot{k}$'s should tend towards $\dot{s}$'s that are lower (higher). This is not to say that the original differences need disappear. Rather, to the extent that the disparities were exaggerated by fluctuations, they should diminish.

Note that the $\beta(\dot{k}-\mu)$ term in (3) allows for the "shrinkage" of the arrest-rate variations. If $0 \leq \beta < 1$, the projected $\dot{s}$'s preserve the ordering of the $\dot{k}$'s but show a smaller degree of divergence.

To be sure, the β-related term is just a first-order model of the underlying regression effect. A more precise representation of the process would require assumptions about the "mixing distribution" of the true Poisson parameters. But, as we will soon see, the simplifying features of (3) seem surprisingly accurate in practice.

The regression phenomenon goes beyond the aspect considered so far; superimposed on the effect just discussed is another quite different in nature. If we only study youths with (say) a certain number of arrests, those apprehended at artificially low rates are more likely than others to be excluded from the sample. Hence those who do make it into the data set will, on average, have $\dot{k}$'s that exceed their λ's (i.e. the asymmetric "censoring" reduces under-estimation more than overestimation).

The upshot of this condition is that, even if all the λ's were stable and retirement from crime were nonexistent, one would expect the average $\dot{s}$ to fall below the average $\dot{k}$. This circumstance can be reflected in (3)'s α-term: if $\alpha < \mu$, a delinquent whose $\dot{k}=\mu$ (i.e. an "average" delinquent) has his $\dot{s}$ estimated by α, which is $\mu-\alpha$ units below his earlier arrest rate.

There are, in summary, two distinct regression processes at work. One is a general correction for inflated $\dot{k}$-values, values that arise when high-rate offenders are identified from actual records. The other, more traditional kind (downward pressure at the top, upward pressure at the bottom), creates non-uniformity in the extent of such inflation: it is typically most severe at the highest $\dot{k}$'s, and weak if not nonexistent at the lowest ones. In (3), α tries to measure the average regression effect for the group, while $\beta(\dot{k}-\mu)$ tailors the adjustments for the individual members.

We calibrated (3) with data about the 164 youths arrested a fifth time between 15 and 16.5. Their $\dot{k}$'s and $\dot{s}$'s were estimated to and from, respectively, the instants of fifth arrests. The μ-value for the group was 1.87. Ordinary-least-squares (OLS) analysis on the data yielded the formula:

$$\dot{s} = \underset{(10.76)}{1.52} + \underset{(8.04)}{.286}(\dot{k} - 1.87) \tag{4}$$

(The numbers in parentheses are t-statistics.)

Because the OLS method ignores the heteroscedasticity in the data, the t-statistics above should not be taken too seriously. But if the general assumptions that underlie (4) are correct, the parameter estimates themselves should be unbiased. The α-value of 1.52--the average of the projected $\dot{s}$'s--is about 19% below the mean $\dot{k}$ of 1.87. The drop reflects both the average regression tendency and also, presumably, a limited amount of true desistance between a_5 and 18. Because β = .286, the estimated dispersion of the $\dot{s}$'s is about 70% less than that existing among the $\dot{k}$'s.

Table 5 sheds light on both the specific implications of (4) and on its reliability. The 164 youths considered, ranked by their $\dot{k}$'s, were broken into successive groups of size 15. (The last group had 14 members.) As the first two columns of the table make clear, the $\dot{s}/\dot{k}$ ratio was far above unity in first few groups and far below in the last few. This is a graphic depiction of classic regression-to-the-mean.

For the table's third column, $\dot{s}$-values were multiplied by times observed to yield projected numbers of further arrests. These numbers, calculated on a youth-by-youth basis, were then summed to obtain the group totals that are presented. The corresponding actual numbers of arrests appear in the fourth column.

TABLE 5: SOME CONSEQUENCES OF USING EQUATION 4

GROUP	MEAN $\dot{k}$	MEAN PROJECTED $\dot{s}$	FUTURE ARRESTS: PROJECTED	FUTURE ARRESTS: ACTUAL
1	.38	1.10	31.0	32
2	.51	1.13	31.7	28
3	.65	1.17	29.9	27
4	.87	1.24	29.4	23
5	1.04	1.28	35.9	42
6	1.16	1.32	39.7	39
7	1.42	1.39	36.8	36
8	1.78	1.50	38.3	31
9	2.42	1.68	46.4	54
10	3.92	2.11	48.2	53
11	6.51	2.85	58.4	59

NOTE: Projected arrests do not grow as much as projected $\dot{s}$-values because the more active offenders, being more prone to detention, have shorter "free periods" to accumulate arrests.

Under the Poisson model, the differences between the table's various expected and actual totals should behave like the usual fluctuations of Poisson variates around their means. (We are invoking the theorem that a sum of independent Poisson variates is itself Poisson.) The last two columns in Table 5 are easily consistent with such fluctuations: when constructed from the expected levels in column three, the 95% confidence intervals for the actual levels contain <u>in every case</u> the adjacent entries in column four. And yet, the actual numbers are not so close to the projected ones as to cast doubt on the Poisson model for that very reason.

In other words, if--as many criminal justice researchers believe--personal arrest data are subject to irreducible Poisson randomness, Equation (4) has performed as well as anyone could reasonably expect. And the Poisson theory, though surely not perfectly right, has received empirical support because $\dot{k}$ had predictive value while k and $\ddot{k}$ did not. Thus a "symbiotic" relationship exists between the results here and the stable Poisson model, in which each serves to increase the credibility of the other.

We should note that all these calculations assume tacitly that detention has no effect on a youth's post-release arrest rate. Were this assumption horrendously wrong, we would not have expected (4) to perform as well as it did. This analysis thus provides at least indirect support for one of the prime tenets of selective incapacitation.

XI. SOME FURTHER NEWS

As the last discussion implies, the strength of the "regression-to-the-mean" tendency depends on the rules by which the youths under study were chosen. Hence a formula calibrated for (say) boys with fifth arrests between 15 and 16.5 would not be expected to apply with great generality. Those in the business of devising predictive formulas must anticipate having to do so on a cluster-by-cluster basis.

But such considerations cannot fully explain the utter inadequacy of (5) for boys arrested a fifth time between 14 and 15. When 69 such youths are broken into quartiles based on their $\dot{k}$-values, the following pattern arises:

MEAN $\dot{k}$	MEAN ACTUAL $\dot{s}$
1.15	1.37
1.66	1.89
2.11	2.08
2.93	1.11

Interestingly, the average $\dot{s}$-value for these boys (1.61) is fairly close to that for their older counterparts in Table 5 (1.52). But now, the highest average $\dot{k}$ is associated with the lowest average $\dot{s}$.

While it is tempting to come up with "burnout" theories, we remind readers that when these data were considered in Table 4, the Spearman correlation of $\dot{k}$ and $\dot{s}$ did not differ significantly from zero. Whatever the case, however, it is hard to reconcile these numbers with a systematic relationship under which $\dot{s}$ increases with $\dot{k}$.

But in evaluating this outcome, we should bear in mind that only about 1/7 of the chronic offenders have their fifth arrests before age 15. And those for whom $a_5 < 14.5$--the most volatile subgroup of chronic offenders--comprise only 5% of the entire group. Even if a small, readily identifiable subset of youths seems unpredictable, this does not imply that forecasts are inappropriate for the far larger population in which they appear reliable.

For many reasons, the criminal justice system deals differently with the youngest offenders than the others. Perhaps we have come upon an empirical argument for exempting young teenagers from selective incapacitation.

We have come to the end of our data analysis. Perhaps we should say a few words about why we thought it valuable to return to the Philadelphia cohort.

XII. WHY PHILADELPHIA?

Since the Philadelphia cohort study, others related to selective incapacitation have been performed. Perhaps the best known of these efforts is the series undertaken at the Rand Corporation (e.g. Chaiken and Chaiken (1982), Greenwood (1982)). Given the existence of more recent work, some might question the relevance of a further analysis of the Philadelphia data.

As we see it, "why Philadelphia?" has three primary answers:

(1) The Philadelphia data concentrate on young offenders.

In crimes of violence as well as property, youths in their mid-to-late teens are among the most active participants. And while the number of teenagers will diminish this decade. it will increase again in the next. (The number of U.S. births rose 16% between 1976 and 1982.) Especially when there is talk of blurring the legal distinction between young and adult offenders, we need more precise understanding of the way delinquents behave.

And the sheer size of the Philadelphia data base allows subtle measurements of delinquent behavior. Given the strong "random component of personal arrest records, the study of (say) juvenile desistance requires hundreds of different youths. In the Philadelphia cohort, they can be found.

(2) This Philadelphia study has been based solely on official police records.

Data arising from other sources are subject to some problems. The Rand studies, for example, were based on the self-reports of prison inmates. While the researchers went to enormous lengths to cut potential abuse, there is just no protection against the shrewd liar who is consistent, and who tells the truth about things that can be checked while distorting shamelessly about those that cannot.

Moreover, some discussions of selective incapacitation presume that, at the time of sentencing decisions, judges will have access to certain information that are not part of official police files. If this assumption is unrealistic, the benefits attributed to selective incapacitation might be overstated.

Obviously, a study based on official records is invulnerable to such difficulties. But it has problems of its own. For one thing, arrests are not convictions; some of them are erroneous and, if not substantiated by judicial proceedings, their use in later sentencing decisions seems questionable. Furthermore, the fraction of crimes that lead to arrests varies in unknown manner from offender to offender. Thus those delinquents who seem the most active might, in reality, be simply the clumsiest. Such a pattern would clearly distort any inferences about the crime-reduction potential of selective incapacitation.

But when different kinds of data suffer from very different flaws, we do better to consider all of them than to favor one kind arbitrarily. And, among official data, the Philadelphia cohort records are an exemplary set.

Another aspect of using official records might relate to fairness. A selective imprisonment policy seems easiest to defend when it is consonant with the "just desserts" theory of punishment. One could argue that an offender's affront to society grows with the number of crimes he commits, with the speed at which he commits them, and with the continuing growth in the frequency of his acts. For this reason, projections of future criminality based on k, $\dot{k}$, or $\ddot{k}$ might offend the sense of justice less than others that depend on "noncriminal" factors, like marital or employment status, or education. (Such variables have been present in other studies.) Through restricting ourselves to the police record, we have been considering how well selective incapacitation can succeed in a framework that, while not identical to the "just desserts" paradigm, does not depart from it abruptly.

(3) The Philadelphia cohort studies are among the best-known work in criminology.

As the *Newsweek* quote implied, their direct and indirect influence is immense. What people perceive the Philadelphia data to be saying is inevitably of importance in a range of policy debates. Any analyses that might increase the depth and detail of these perceptions would seem worthwhile.

In addition, a more recent Philadelphia cohort has been followed: those youths of both sexes born in 1958. The various findings here--and the methods used to obtain them--could be of value in the study of the new cohort.

FINAL REMARKS

Our studies of the Philadelphia cohort data suggest that among older, active delinquents, there is a simple montonic relationship between past and future patterns of arrest. The results thus imply that, contrary to some fears expressed earlier, predicting future delinquency need not be futile, nor need it depend on variables the use of which seems reprehensible (e.g. race).

We have not dealt directly with the issue of "false positives" (i.e. those youths with bad prognoses who in reality abstain from further crime). But the findings on false desistance suggest that the problem might arise only rarely. Moreover, harsh punishment for those with the k's and $\dot{k}$'s could be defended as a just response to the record that already exists.

It would exaggerate matters, however, to say that we have made a clear case for selective incapacitation. As noted, arrest rates need not be proportional to the underlying crime commission rates. But even neglecting this, there is the point made in Section IV: once we discard first and second offenders, arrest rates do not really vary that much among Philadelphia delinquents. Lacking a subset of offenders far worse than their "competitors" for prison space, one has trouble constructing a "practical" argument for focusing imprisonment on them.

Hence our earlier characterization of the results as "bittersweet." If they do not confirm the worst fears of the selective incapacitation advocates, neither do they support their best hopes.

REFERENCES

Avi-Itzhak, B. and R. Shinnar (1973), "Quantitative Models in Crime Control," Journal of Criminal Justice, 1(3).

Barnett, A. (1981), "The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies," Operations Research, 29(2).

Barnett, A. (1982), "An Underestimated Threat to Multiple Regression Analyses Used in Job Discrimination Cases," Industrial Relations Law Journal, 5(1).

Blumstein, A. and S. Moitra (1980), "The Identification of 'Career Criminals' from 'Chronic Offenders' in a Cohort," Law and Policy Quarterly, 2(3).

Blumstein, A. and D. Nagin (1978), "On the Optimum Use of Incarceration for Crime Control," Operations Research, 26(3).

Chaiken, J. and M. Chaiken (1982), Varieties of Criminal Behavior, RAND report R-2814-NIJ.

Chaiken J. and J. Rolph (1980), "Selective Incapacitation Strategies Based on Estimated Crime Rates," Operations Research 28(6).

Clarke, S. (1974), "Getting 'em Out of Circulation: Does Incarceration of Juvenile Offenders Reduce Crime? " Journal of Criminal Law and Criminology, 65(4).

Figlio, R. (1981), "Delinquency Careers as a Simple Markov Process," Chapter 2 in Models in Quantitative Criminology (J. Fox, Editor), Academic Press, New York.

Greenwood, P. (1982), Selective Incapacitation, RAND report R-2815-NIJ.

Maltz, M. and S. Pollock (1980), "Artifical Inflation of a Delinquency Rate by a Selection Artifact," Operations Research, 28(3).

Sellin, T. and M. Wolfgang (1964), The Measurement of Delinquency, Wiley & Sons, New York.

Tierney, L. (1983), "A Selection Artifact in Delinquency Data Revisited," Operations Research, 31(5).

Von Hirsch, A. (1976), Doing Justice: The Choice of Punishments, Hill and Wang, New York.

Wolfgang, M., T. Sellin, and R. Figlio (1972), Delinquency in a Birth Cohort, University of Chicago Press, Chicago.

APPENDIX A: CRIME SERIOUSNESS AS A FUNCTION OF TOTAL JUVENILE ARRESTS

Throughout this paper, we neglected the curfew violations recorded in juvenile arrest files. The curfew laws were eccentric to Philadelphia and have since been abolished there; their reenactment seems an exceedingly remote prospect. Thus to tie a predictive scheme to the curfew law seems no more practical than to do so for the Prohibition Act.

Once curfew violations were excluded, all other offenses were classified as "serious" or "minor" in the preparation of Table 1. The serious crimes, as noted in the text, are the FBI Index Offenses plus arson, kidnapping, and malicious mischief. Let P_k be the fraction of arrests that were serious among youths whose record lengths were k. P_k was calculated from the formula:

$$P_k = \frac{V_k}{kC_k}$$

where C_k = number of boys with exactly k arrests

V_k = total number of serious arrests those C_k youths sustained

To make Table 1 easier to read, we generally combined P_k's for two successive values of k. If these two values were t and t+1, the statistic presented was $\alpha_t P_t + (1-\alpha_t)P_{t+1}$, where $\alpha_t = C_t/(C_t + C_{t+1})$. Hence we simply weighted P_t and P_{t+1} in proportion to the number of boys with each record length.

To test the "robustness" of the Table 1 pattern to changing definitions of seriousness, we performed an experiment. We reclassified petty larceny and malicious mischief as minor offenses, and designated the remainder of the serious crimes as "very serious." Table 1A presents statistics based on the new dichotomy.

TABLE 1A: CRIME SERIOUSNESS VS. TOTAL RECORD LENGTH UNDER A MORE RESTRICTIVE DEFINITION OF "SERIOUS"

TOTAL NUMBER OF OFFENSES	PERCENTAGE THAT WERE VERY SERIOUS
3-4	31.4
5-6	33.3
7-8	35.9
9-10	37.5
11-12	28.7
13+	35.8

The new rules increase the proportion of minor offenses from 48% to 66%. But the lack of association between record length and crime seriousness is as conspicuous as before.

Had we redefined "serious" still more restrictively, the pattern might have changed slightly. Wolfgang et al report that the chronic offenders, who absorbed 52% of the entire cohort's arrests, were responsible for 2/3 of the arrests for violence. But so long as crime seriousness does not <u>decline</u> with record length, searching for the frequent offenders clearly is related to searching for the worst ones. That relationship does not seem in jeopardy in the Philadelphia data.

APPENDIX B: CORRECTING FOR IMPRISONMENT

Many calculations in this paper depend on estimating s_k, a delinquent's arrest rate for just after a_k, his age at k^{th} arrest, to age 18. Getting an estimate becomes tricky if the youth was in detention over part of that period.

We will proceed on the assumption that, except while confined, the youth accumulates arrests under a Poisson process with constant mean. Empirical support for this approximation appears in the main text beginning in Section IV; the relevant evidence is summarized in Section X.

We also assume that confinement decisions were not tied closely to the offender's arrest rate just before they were imposed. If this axiom is wrong, there is a danger of estimation biases of the kind discussed by Maltz and Pollock (1980) and Tierney (1983). We readily concede that the assumption is imperfect; however, given that we consider only multiple offenders, and given the kinds of cross-group comparisons we are making, we believe the problem is of secondary importance.

If a youth is not arrested at all between a_5 and 18, a natural expression for $\dot{s}_k$ is given by:

$$\dot{s}_k = \frac{n_k}{18-a_k} \tag{B-1}$$

where n_k = number of juvenile bookings <u>after</u> the k^{th}

(B-1) is easily shown to yield an unbiased estimate of the underlying Poisson parameter.

If a delinquent received some detention in consequence of his m^{th} arrest, the code number 5 appears in the data file after a_m. But no information is given about the <u>length</u> of his period of confinement. After consulting with

Sobelevitch, Director of Youth Services of the Pennsylvania Department of Corrections, Clarke (1974) estimated that each detention was for approximately 9 months.

Our own conversations with Sobelevitch informed us that the variation of "sentences" around their mean value was exceptionally high. Many delinquents were sent to the Penny Pack Facility, where they experienced nominal confinement on the order of two weeks. But others were dispatched to Glen Mills or Camp Hill, where stays averaged 15 to 18 months. The code-5 designation does not distinguish these vastly different situations.

Given these circumstances, our inclination was to delete from the time line those intervals that were "contaminated" by unknown periods of detention. More precisely, suppose a delinquent had one "code 5"* between a_k and 18, at age a_m.

Then:

(i) If he was subsequently arrested at a_{m+1}, we estimated $\dot{s}_k$ by:

$$\dot{s}_k = \frac{n_k - 1}{18 - a_k - (a_{m+1} - a_m)} \tag{B-2}$$

(ii) If his last arrest was at a_m and, in addition, $a_m > 16$, then the estimate was:

$$\dot{s}_k = \frac{n_k - 1}{a_m - a_k} = \frac{m-k-1}{a_m - a_k} \tag{B-3}$$

Both these estimators are unbiased under the Poisson model. (See Appendix C for some discussion of (B-3).)

* If there were two or more detentions, we generalized these procedures.

Suppose it were known that a given delinquent experienced one detention of length exactly L. Even if that confinement was caused by some arrest after a_k, an unbiased estimator of his Poisson parameter can be shown to follow:

$$\dot{s}_k = \frac{n_k}{18-a_k-L} \qquad \text{(B-4)}$$

In a few situations, we thought it preferable to use an approximation based on (B-4) rather than on (B-2) or (B-3). Two examples help identify such cases.

(iii) Suppose a_k was "close" to 18, a_m was (say) 17.3, and a_{m+1} = 17.6. Given that $a_{m+1}-a_m$ = .3, it is extremely probable that the delinquent went to Penny Pack rather than Glen Mills or Camp Hill; thus he was "out of circulation" for about two weeks (.04 years). We would approximate his $\dot{s}_k$ by $n_k/(18-a_k-.04)$.

(iv) Suppose a_m = 14.6 and a_{m+1} = 17.5. Even if the youth got a long sentence at a_m (around 16 months), there was clearly a long interval in which he was "on the street" but not arrested. Rather than disregard this information, we would estimate his $\dot{s}_k$ by $n_k/(18-a_k-1.33)$.

In other words, we would estimate the period of detention rather than delete the headway containing it if (a) the uncertainty in the time served was less important than usual <u>and</u> (b) with the headway removed, the time remaining for measuring $\dot{s}_k$ was rather short.

APPENDIX C: FALSE DESISTANCE

In the text, we outlined a test of the hypothesis that all observed desistance after the fifth arrest was false. The test assumed that youths are arrested under stationary Poisson processes and that, therefore, a boy whose rate parameter is λ has a probability $\exp(-\lambda[18-a_5])$ of desisting falsely. (The probability is different if the fifth arrest led to confinement.)

We obviously need an estimate of a youth's λ-value given the dates of his first five arrests. An intuitively appealing estimator, $\tilde{\lambda}$, follows:

$$\tilde{\lambda} = \frac{k-1}{a_k-a_1}$$

The motivation for $\tilde{\lambda}$ is that the arrest at a_1 triggers our awareness of criminal activity that has been going on for some unknown amount of time. We do know that k-1 additional arrests occurred in the next a_k-a_1 time units; $\tilde{\lambda}$ takes the quotient of these two numbers for a rate per unit time.

But $\tilde{\lambda}$ is subject to considerable bias. If a youth is observed continuously, his ratio of arrests to time free should oscillate randomly around λ, with the peaks of the oscillatory cycle being the instants of arrests. But it is precisely at one of these peaks that $\tilde{\lambda}$ is calculated; hence it suffers an upward distortion.

Fortunately, the problem just cited is easy to correct. One can show that, ordinarily, the upward bias would disappear with the adjusted estimator λ^* that follows:

$$\lambda^* = \frac{k-2}{a_k-a_1}$$

We would have proceeded with λ^* for k=5 were it not for a quirk in the data. Examining the arrest records of various multiple offenders, we found a substantial fraction with one isolated arrest several years prior to evidence of sustained criminal activity. Thus, to estimate reliably their arrest rates once youths reached "cruising speed," we adopted a convention: we would start observation at the time of the second arrest (a_2). Treating the arrest at a_5 as the fourth in the period of study, we approximated the Poisson parameter by $\hat{\lambda}$ given by:

$$\hat{\lambda} = \frac{2}{a_5 - a_2}$$

Note that $\hat{\lambda}$ is analogous to λ^* for k=4 and that, even if excluding the initial arrest was not necessary, doing so should not create bias in $\hat{\lambda}$.

Unfortunately, two further complications arise in the situation at hand. Even if $\hat{\lambda}$ were unbiased, it should overestimate the λ's for roughly half the youths (Group 1), and underestimate for the rest (Group 2). But, in our hypothesis test, we only consider boys with fifth arrests prior to age 18. This restriction should exclude more youths in Group 2 than Group 1; hence a positive bias in the estimates of λ has not been fully expunged.

On the other hand, we use the λ-estimates in the exponential expression for the chance of no arrests. Even an unbiased estimator of λ would not yield an unbiased estimate of the relevant probability; this is a special case of the theorem that the mean of a function of a random variate need not equal that function at the variate's mean. The tendency would be to get desistance probabilities that are artificially high.

In other words, the overestimation of λ by $\hat{\lambda}$ serves to dampen an upward distortion in the use of $\tilde{\lambda}$. While it is too much to hope that the two biases cancel each other, rigorously correcting for them would be very difficult. For an order-of-magnitude estimate of the level of false desistance, the trouble would seem greater than it is worth.

Using the procedure involving Q_i's that was described in Section we calculated that--barring any true desistance--the number of Table 3 youths with no arrests after the fifth would have mean 26.6 and variance 22.3. Hence, the quantity $\mu+1.64\sigma$, the accept/reject threshold in a one-sided test of the null hypothesis (5% level), would be 34.3. At 35, the actual number of desistors falls at the lower edge of the rejection region.

That true desistance is altogether absent hardly seems credible; thus we have no inclination to quibble about these marginal test results. What is noteworthy is that true "retirement" might be so infrequent that its very existence is hard to verify.