Center for Research in Criminal Justice

University of Illinois at Chicago Circle
Box 4348 Chicago 60680

A Reanalysis of UDIS: Deinstitutionalizing the Chronic Juvenile Offender*

Richard McCleary
Arizona State University

Andrew C. Gordon
David McDowall
Northwestern University

Michael D. Maltz
University of Illinois, Chicago Circle

November 1978

*Work on this report was funded in part by a contract from the Illinois Department of Corrections to the Center for Research in Criminal Justice of the University of Illinois, Chicago Circle.

Contents

I. Introduction:

The purpose of this paper is to provide a detailed re-examination of UDIS: Deinstitutionalizing the Chronic Juvenile Offender, prepared for the Illinois Law Enforcement Commission by Charles A. Murray, et al of the American Institutes for Research (AIR). The AIR evaluation is of a program called "Unified Delinquency Intervention Services" (UDIS), a set of alternatives for Cook County juvenile offenders who would otherwise be incarcerated in the Department of Correction (DOC) for delinquency. Murray and his associates report a profound decrease in the delinquent activity of youths placed in either UDIS or DOC, and no substantial difference between the programs in the post-release delinquent behavior of the youth. In both cases, rapidly accelerating rates of delinquent activity are reported to be drastically reduced (the "suppression effect," in the terminology of the AIR report) after youths were "treated" by either of the two programs. The authors further conclude that the differences in between UDIS and DOC were slight; that no less severe prior intervention (e.g. station adjustments, probation) had any salutary effect on these serious delinquents; and that the more intense the UDIS placement (e.g. wilderness programs and intensive psychological treatment, compared to advocacy and educational programs), the greater the impact on recidivism.

This suppression effect, they point out (p. 193), was also noted in two prior studies of juvenile delinquency treatment programs (Empey and Lubeck, 1971; Empey and Erickson, 1972). Both of these studies reported a decrease in delinquent activity of about 80 percent in comparing the twelve months before intervention to the twelve months after intervention. Since these studies were based on controlled experiments, data were available for both experimental and control groups; the decrease was about the same for experimentals and controls. This finding had been accepted by the research community without much comment, and gave support to the existence of a suppression effect in the UDIS study.

We demonstrate in this paper that, after a reanalysis sensitive to potential artifacts in the UDIS data, there is no evidence for a suppression effect attributable to the correctional interventions. In addition, although we have not analyzed the data in the other two studies, the before - after decrease in delinquency can probably be explained by the regression artifact described below.

Most studies of juvenile and adult offenders measure recidivism in a gross fashion: if an offender is even once re-arrested* he is considered to have recidivated, equivalent to an offender who committed multiple offenses after program release. The AIR study substitutes a far more meaningful measure: rates of delinquent activity, or the proportion of offenses reduced in a pre-/post-intervention comparison of delinquent careers. But the means of calculating the rates of delinquency served to artifically enhance the suppression effect, and to obscure the impact of artifacts from all but the most careful statistical scrutiny.

The AIR authors also note in passing that some alternative explanations (statistical artifacts) may have a minimal influence on their findings, but none of any importance. Our reanalysis measures the impact of these statistical artifacts, notably regression (a misleading and purely statistical drop in delinquency due solely to selection of youths at a time of abnormally high pre-intervention rates) and maturation (the acknowledged tendency for youths

* or re-committed, or re-appeared in court, etc., whatever the index of criminal activity

to commit fewer offenses as they grow older). Using more advanced analytic techniques, we can explain how these two artifacts probably account for the bulk of the suppression effect. In addition, we discuss in detail the clear presence of another artifact not discussed in the AIR report: mortality (or the increasingly biased loss of delinquents in pre-/post-comparisons of rates.)

Since so much of the AIR report and nearly all the widespread attention following its publication has been devoted to the apparent "suppression effect," that effect will be the major focus of this paper. Our re-analyses have been conducted using a data tape obtained from AIR. Our reanalysis is organized in the following way. Section 1 summarizes our findings and discusses the effect of the research design chosen on the validity of AIR's results. Section 2 reviews the AIR findings and the part played by the mortality, maturation and regression artifacts. Section 3 treats UDIS vs. DOC. Section 4 discusses the AIR design with respect to within-UDIS contrasts, and Section 5 is a technical appendix.

We then comment briefly on how, given the nature and intent of UDIS, the AIR design provided an inappropriate perspective on the program they evaluated, and present some results from an alternative analysis of AIR's UDIS data. We should point that we have been limited in our reexamination to existing data. We had neither the ability nor the resources to collect new data.

Research Design

In the broadest sense, there are three types of research designs: true experimental, quasi-experimental, and correlational. From an analytic perspective the best designs by far are true experiments, which require random assignment of subjects to groups (e.g. UDIS or DOC), and the presence of a no-treatment control group (e.g. no correctional intervention). True experiments have the virtue of controlling all "threats to internal validity," so that only the intervention remains as a plausible explanation for any obtained differences in outcome between groups.

But true experiments are not always feasible in delinquency research. Few judges would permit an apparently serious delinquent to remain untreated merely for the sake of scientific integrity.**

Correlational designs are analytically weak, and ordinarily require vast sample sizes. Thus they are not widely used in the evaluation of delinquency intervention programs.

The AIR evaluation of UDIS employed a quasi-experimental design which has been called a "pre-to-post single-group comparison." In the terminology of Campbell and Stanley (1963) this can be diagrammed as

0	X	0
pre		post,

where the <u>O</u>s denote measurement (e.g. rates of delinquent behavior), and the <u>X</u> denotes the intervention, (e.g. UDIS or DOC). While there were three groups of delinquent (UDIS, DOC, and a Pre-UDIS DOC baseline), with before and after measures, the design cannot afford the randomization a true experiment would require, and is best analyzed as three separate, single-group quasi-experiments.

A quasi-experiment cannot control for threats to internal validity, and thus plausible alternative explanations for an obtained difference between groups must be considered in detail. In the next section we will see why this is a crucial distinction when evaluating the AIR data.

Given the quality of the recidivism data available to AIR, and the

**Notable exceptions are Empey & Lubeck (1971) and Empey & Erickson (1972).

nature of the evaluation task*** their choice of a quasi-experimental design was optimal. But while there are no inherent controls for internal validity in a quasi-experimental design, a researcher can often cultivate supplemental data or strategic analytic techniques to minimize the impact of these plausible rival explanations. Since we have had no opportunity to collect supplemental data, we have relied on analytic techniques to account for these threats to validity.

*** We refer here to the evaluation task as defined by AIR. Particularly within the context of UDIS, the evaluation is distorted by an extreme emphasis on recidivism. We shall address this point at the end of our report.

1.1 Introduction: A Summary of Our Findings

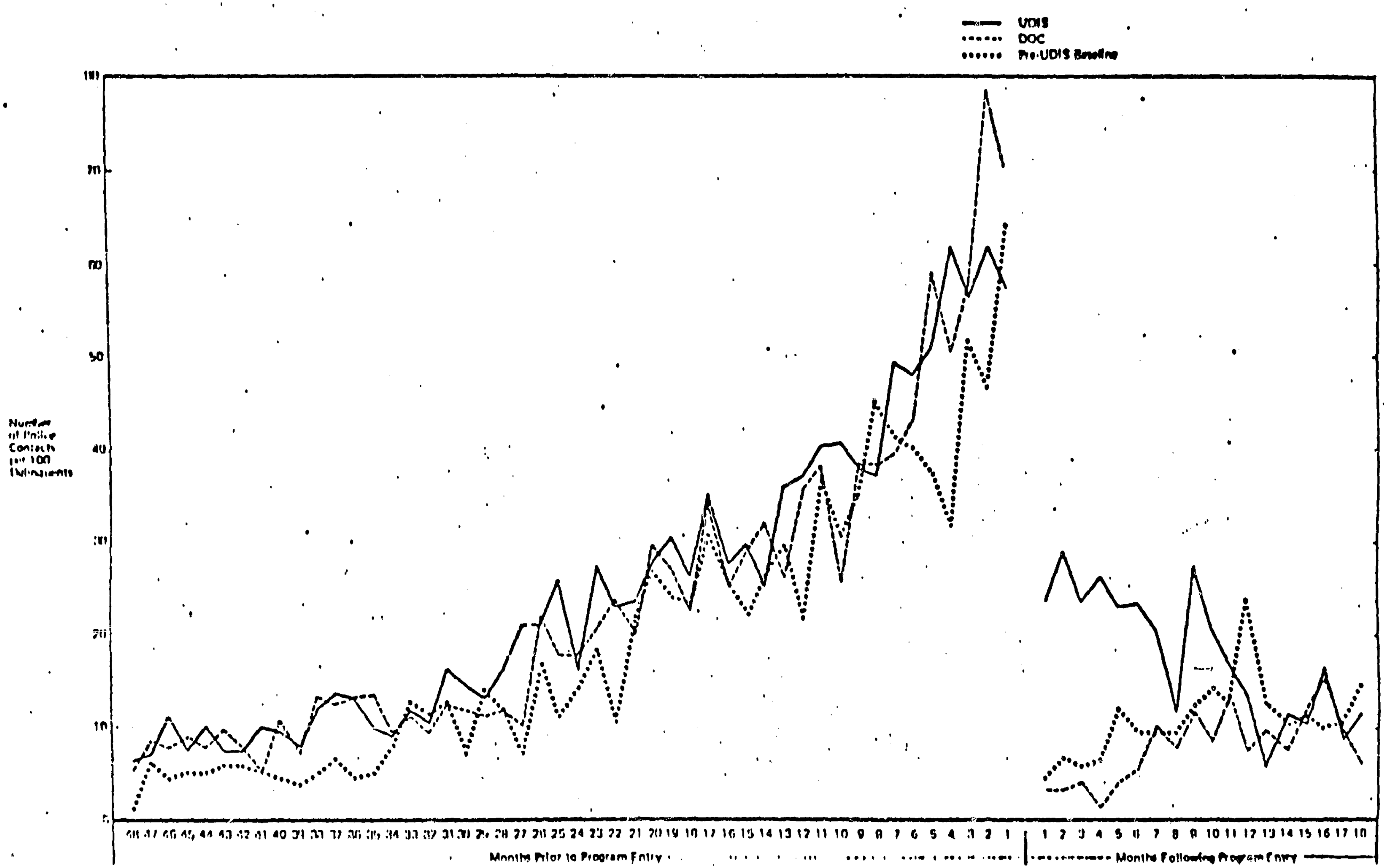
In Figure 1.1 on the following page, we have reproduced the most controversial "finding" of the AIR report (their Figure III.1). This is a graphic depiction of the "suppression effect" reported by AIR. All three groups, UDIS, DOC, and Pre-UDIS Baseline, show drastic reductions in arrest rates from pre-intervention to post-release. This figure is an illusion, however. From a reanalysis of the AIR data we have found that:

(1) The "suppression effect" is enhanced by experimental mortality during the post-release period. There are 103 "missing" delinquents, that is, delinquents who appear in the pre-intervention statistic but not in the post-release statistic. Throughout the post-release period, delinquents tend to "disappear" either because the length of follow-up was artifically shortened or because the delinquents committed new crimes and were returned to institutions or programs. We have found that these "missing" and "disappearing" delinquents are the "worst" or "most active" delinquents on the average. Mortality in this case inflates the AIR estimate of "suppression."

(2) The "suppression effect" is enhanced by maturation. It is widely accepted that the individual's rate of delinquency changes with age. We have found that the older the delinquent at the time of release from institution or program, the greater the reduction in post-release arrests; the greater the "suppression effect." Maturation is a particularly potent artifact because these delinquents have spent nine months on the average in institutions or programs. During this period of time, arrest rates are expected to drop regardless of any "suppression effect."

(3) The basis of the "suppression effect" reported by AIR is a regression artifact. These delinquents have been selected from the population of all delinquents eligible for a correctional intervention. Moreover, in each case

Figure 1.1: Taken from Murray, Thompson, and Israel (1978), shows three groups of delinquents 48 months before and eighteen months after a "radical" intervention



the _time_ of intervention has been selected. If juvenile judges bias their selection to favor those delinquents who appear to be the "worst" of all, data similar to those in Figure 1.1 are generated. _Using only Number of Pre-intervention Crimes_ (which is correlated with the regression artifact) and _Age at Release_ as independent variables, we are able to explain nearly half the variance in the _Suppression_ scores for a sample of delinquents. This finding supports our claim that the AIR "suppression effect" is nothing more than a methodological artifact.

(4) The AIR report explains the "suppression effect" in terms of a deterrence mechanism. However, we find a negative relationship between _Time Served_ for the incident offense and _Suppression_. The shorter the length of time spent in institutions or programs, the greater the "suppression" effect. This finding counters the AIR deterrence rationale.

(5) The AIR report finds no substantial difference between UDIS and DOC with respect to recidivism. Although we used a more powerful statistical model which identified slight differences between UDIS and DOC, the differentials were not statistically significant. On this point our conclusion coincides with that of AIR.

(6) The AIR report did not contrast UDIS and DOC with Pre-UDIS Baseline. Our analysis of this contrast found that UDIS and DOC have much lower real recidivism rates than Pre-UDIS Baseline. The difference is both statistically and substantively significant. While there are a number of interpretations of this finding, there is no doubt recidivism rates in Illinois have dropped since the introduction of UDIS. (See Section 3.0).

2.0 The Major Finding: A Suppression Effect

For all practical purposes, we can say that the AIR report has a major finding and a minor finding. In this section, we address the major finding: a suppression effect.

On page 163 of the report, the AIR authors discuss the results of a before/after regression. They note that

> "The difference between the levels of pre and post intervention activity are striking. During the three months preceding intervention, for example, the average number of police contacts per 100 (delinquents) per month was 60.5. During the three months following release from the intervention program, the average number was 17.7 -- a reduction of 71 percent. During the half year "before," the average number was 54.8; during the first half year "after," the average was 18.6 -- a drop of two-thirds. The reductions continued throughout the period of postprogram observation.
>
> "We label the phenomenon the 'suppression effect.' Delinquent activity did not stop altogether. Evidence is not invoked that any of the youth was rehabilitated or inspired to go straight, or that any other change in the youth's internal state took place. Behaviorally, a certain type of activity appears to have been suppressed to a point far below its level prior to intervention, for whatever reasons.
>
> "Reductions of better than two-thirds are substantial. They are, in fact, material for headlines...."

The suppression effect finding has indeed become "material for headlines." Reports of the major finding have appeared in the Chicago Tribune and Sun-Times as well as the New York Times; in corrections journals such as the Criminal Justice Newsletter and Federal Probation; and the senior author, Dr. Charles Murray, has described the major finding in a statement to the U.S. Senate Subcommittee on Juvenile Delinquency.

Until now, however, the major finding has not been subjected to the scrutiny of the social science community. In what follows, we report the results of a reanalysis of a data tape purchased from AIR. Our reanalysis suggests unambiguously that the AIR major finding has been blown out of proportion. While the AIR report and popular media have consistently cited a suppression effect on the order of "two-thirds reduction," we have been unable to find any evidence whatsoever of a suppression effect. We have concluded instead that the AIR report of a suppression effect is due to methodological artifacts.

What has passed for a suppression effect is in our opinion due instead to three threats to internal validity: mortality, maturation, and regression. A threat to internal validity means generally that some facet of the research design has confounded the analysis, making some treatment appear effective. In this case, a radical correctional intervention is made to look effective by certain flaws in the research design. Had alternative designs been used, the AIR findings would have been different.

(1) Mortality refers generally to a selective attrition of cases from the treated group. In the AIR report, the analysis assumes that attrition is essentially random. It is not. The more active or more serious delinquents are held in programs and institutions for longer periods of time on the average than the less active or less serious delinquents. As a result, the post-release measures of delinquency which are the basis of the suppression effect tend to concentrate on the least delinquent youths. We have found that 103 delinquents (over 21% of the sample) were not observed at all in the post-release period. We call these 103 the "missing" delinquents, and on the average, these 103 "missing" delinquents are the most active of all. By including their behaviors in the pre-intervention measures but excluding their behaviors in the post-release measures of delinquency, the AIR report has drastically overstated the magnitude of the suppression effect.

Unfortunately, the data tape made available to us by AIR does not include the pre-intervention behaviors of these 103 "missing" delinquents. We cannot therefore make a precise estimate of mortality artifact. We can only say that it is substantial. Our analysis of mortality in the AIR major finding is restricted to an analysis of the relationship between seriousness of the pre-intervention delinquent career and mortality and to an analysis of the arithmetic procedures used in the AIR report analyses. There are many acceptable arithmetic procedures for computing a rate of delinquency. The method used in the AIR analyses interacts with mortality to further exaggerate the impact of correctional interventions.

(2) Maturation refers to a change in behavior that is due to nothing more than a natural aging process. As a youth matures, his rate of delinquency will generally increase, but then at some point his rate of delinquency will begin to decrease. There are a number of plausible reasons for this decrease. Some youths find jobs, for example, and the time spent at work keeps these youths away from the influence of delinquent peers. Gainful employment also removes the stimulus for instrumental delinquent acts. There may also be a deterrent effect due to nothing more than age. As the delinquent matures, his inappropriate behavior is more and more likely to result in criminal court sanctions rather than in the less severe juvenile court sanctions. Whatever the reason, prior research has consistently pointed to the effects of maturation on delinquent activity and our reanalysis of the AIR data suggests that a substantial portion of the suppression effect is due to nothing more than maturation.

In the AIR report itself, a cursory analysis gave clear and unambiguous evidence of a maturation effect. However, the AIR authors conclude that the maturation effect is "small." After explaining the problems with the AIR analysis for maturation, we reanalyze the data and conclude that the effect is large. As delinquents are held in programs and institutions for relatively long periods of time, the effects of maturation comprise a greater and greater portion of the suppression effect. However, the method in which the AIR report aggregated the delinquent careers obscures the effect of maturation. We have disaggregated these data so that the effect of maturation is not obscured.

(3) As an appendix to this report, we include a paper by Maltz and Pollock which lays the mathematical foundation for a regression artifact. When analyzing delinquency rates, it is important to remember that the delinquents have been selected for intervention. Maltz and Pollock show that, when this is the case, a suppression effect is guaranteed regardless of how effective the correctional intervention may be. The Maltz and Pollock paper is rather technical and not easy reading for the general audience. However, the mathematical principles which they outline form the basis of our reanalysis.

Most of the research on regression artifacts concerns only paper-and-pencil testing situations. Students who score extremely high on an achievement test, for example, will ordinarily score lower on a second test. Likewise, students who score extremely low on the first test will ordinarily score higher on the second test. This phenomenon is called "regression to the mean." The same phenomenon applies to the evaluation of a delinquency prevention program. However, unlike the paper-and-pencil testing situation, there are no acceptable methods for controlling this regression artifact.

In the AIR report, an inappropriate and weak analysis gave distinct evidence of a regression artifact in the major finding. We have shown how this regression artifact was probably generated. The results of this analysis suggest quite clearly that the regression artifact contributes substantially to the suppression effect.

2.1 Mortality

Mortality refers to a biased attrition from the samples. To illustrate how this threat to internal validity affects the AIR findings, consider four hypothetical delinquent careers:

$$\ldots\; O_{48}\;\; O_{47}\; \ldots\; O_{1} \quad (I\ldots\ldots) \quad O_{1} \qquad\qquad\qquad\qquad\qquad O_{i} \quad 12/31/76$$

$$\ldots\; O_{48}\;\; O_{47}\; \ldots\; O_{1} \quad (I\ldots\ldots\ldots\ldots) \quad O_{1}\; \ldots \quad O_{j} \quad 12/31/76$$

$$\ldots\; O_{48} O_{47}\; \ldots\; O_{1} \quad (I\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 12/31/76$$

$$\ldots\; O_{48}\; O_{47}\; \ldots\; O_{1} \quad (I\ldots\ldots\ldots)\; O_{1}\; \ldots\; O_{k}(I\ldots\ldots\ldots\ldots\ldots\ldots \quad 12/31/76$$

These careers are diagrammed in real time. This is an important point. All of the careers end on December 31, 1976 when the AIR study ended. At that time, some delinquents had been free for a relatively long period. Other delinquents, largely in the UDIS and DOC samples, had been free for only a few weeks or months. Some delinquents had not been free at all, and finally, some had been released but had been sent back as a result of new offenses.

To estimate the suppression effect, the AIR authors aggregated 487 real-time careers into an artificial "phase" time. This is diagrammed as

$$\ldots\; \bar{O}_{48}\, \bar{O}_{47}\; \ldots\; \bar{O}_{1} \quad \text{(Intervention)} \quad \bar{O}_{1}\;\; \bar{O}_{2} \quad \ldots \quad \bar{O}_{i}\; \ldots$$

where each monthly observation is a mean or average rate of delinquent activity for the samples. But while all 48 pre-intervention observations are based on the total combined sample of 487 delinquents, the post-release observations are based on few and fewer observations. First, there are

103 "missing" delinquents, 21% of the sample. These delinquents appeared in the pre-intervention observations but not in the post-intervention observations. Second, from post-release month to month, many delinquents "disappear," the percentage increasing each month.

If this attrition were random, there would be no problem of interpretation. The data suggest unambiguously, however, that the attrition is biased. In Figure 10.5, page 171 of the AIR report, the suppression effect is broken down into three categories of pre-intervention "seriousness." The most serious delinquents have had 17 or more prior police contacts; the medium delinquents have had nine to 16 prior police contacts; and the least serious delinquents have had eight or fewer prior police contacts. The sample sizes and attrition rates by this breakdown are

	Number Pre-	Number "Missing"	Number "Disappeared" by One-Year Post-release
Most	107	43 or 40%	82 or 77%
Medium	247	64 or 26%	169 or 68%
Least	123	31 or 25%	81 or 66%

It is clear from a simple examination of Figure 10.5 that the most delinquent youths are also the most likely to be "missing" at the start of the post-release period; and those most delinquent youths who are not already "missing" are the most likely to "disappear" at any point in the post-release period.

"Missing" and "disappearing" delinquents on the average are more serious or more active than the remaining delinquents. Mortality thus has the effect of reducing the post-release delinquency rate of the group and of inflating the estimate of suppression. Moreover, while these more serious delinquents constitute a relatively small proportion of the combined samples (approximately 20%), they may account for a relatively large proportion of the police contacts for the combined samples. (We have to hedge our language here because our analysis is based only on a subset of the AIR data; we do not have any data on the 103 "missing" delinquents.)

To analyze the effects of mortality, we have tried to predict the "disappearance" of delinquents from the post-release measures. All the variables which we examined and which are ordinarily associated with the seriousness of a delinquent career proved to be good predictors of "disappearance." These variables include:

(1) Age at first arrest is often cited as the single best predictor of post-release success for delinquents. The older the delinquent at the time of first arrest, the less serious the delinquent career and the more likely that the delinquent will succeed after a treatment. Using length of follow-up as the dependent variable, the regression relationship is

Follow-Up = 152.27 days + 1.42 (Age at first arrest)

All regression co-efficients are significant at the 95% level. This analysis shows that the older the delinquent at first arrest, the longer the follow-up. In other words, delinquents who were first arrested at an early age are the most likely to "disappear."

(2) Number of pre-intervention arrests is clearly related to the seriousness or level of activity of the career. The more pre-intervention arrests, the more serious the career. Again, using length of follow-up as the dependent variable, this regression equation is

Follow-up = 462.68 days - 8.5 (Number of Pre-intervention arrests)

Again, all coefficients are statistically significant. This analysis shows that youths who had many pre-intervention arrests were more likely to "disappear" than youths who had few pre-intervention arrests. In other words, the more active delinquents were the most likely to "disappear."

(3) Time in program or institution should be related to the seriousness of the pre-intervention career. Judges presumably give longer sentences to the serious delinquents when sentencing is at issue; institutional personnel presumably keep the more serious delinquents in program or institution for longer periods of time; and so forth. The regression equation for this variable is

Follow-up = 421.64 days - 7.01 (months in program or institution)

All coefficients are statistically significant. This analysis shows that youths who spent longer periods of time in program or institutions were more likely to "disappear" than youths who spent shorter periods of time in program or institutions.

If we consider all three variables simultaneously, the regression equation is

Follow-up = 462.89 days + .42 (age at first arrest)
- 7.9 (Number of Pre-intervention arrests)
- 7.4 (Months in program or institution)

So all the relationships that hold individually also hold simultaneously.

To summarize our analysis of mortality, we can say that a delinquent's chances of "disappearing" are directly proportional to seriousness of the delinquent's pre-intervention career. In our analysis, we have not found even one variable that contradicts this judgment.

UNITED STATES DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
NATIONAL INSTITUTE OF LAW ENFORCEMENT AND CRIMINAL JUSTICE
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE

WASHINGTON, D.C. 20531

9/27/79

Center for Research in Criminal Justice
University of Illinois
Chicago, IL 60680

Please reply to:
NCJRS
Acquisition Report Dept.
Box 6000
Rockville, MD 20850

Dear Colleague:

The National Criminal Justice Reference Service (NCJRS) is an international clearinghouse serving the law enforcement and criminal justice community with a wide variety of information services. In support of these services, we request that you forward a free copy of the following publication(s) for possible inclusion in our bibliographic data base:

"A REANALYSIS OF UDIS."

If a gratis copy is not available, please advise on the sale price so that we might prepare another order. Please do not bill us directly for any item.

Thank you for your courtesy and cooperation in this matter.

Sincerely,

Shu-Shun Chiang
Supervisor of Acquisition

P.S. If you are currently not an NCJRS user and wish to receive further information, please contact NCJRS, User Services Department, Box 6000, Rockville, MD. 20850.

With respect to mortality, we can only say, based on our limited data, that mortality was substantial and that it favored the finding of a suppression effect. The average length of follow-up for our N = 365 delinquents was 357.6 days, slightly less than one year; when 103 "missing" delinquents are also considered, the length of follow-up was only 268 days, slightly less than nine months. But the distribution of follow-ups is skewed in favor of the least delinquent youths and this inflates the suppression effect substantially. As we have no data on the "missing" delinquents, however, we cannot offer a proportional figure.

In the AIR report, troublemakers routinely "disappeared" from post-release samples. The AIR report failed to deal with this threat to internal validity. Whereas a substantial drop in the post-release delinquency rate of the group was due to nothing more than the "disappearance" of the group's most active members, the AIR report incorrectly attributed this drop to the impact of a correctional intervention.

Before we go on to consider the effects of maturation and regression, we will address the arithmetic choice that the AIR authors made in this analysis of the major finding. While mortality alone is a potent threat to internal validity, its potency is magnified in the AIR report by the arithmetic used to estimate delinquency rates.

2.1.1 A Mortality Interaction with Arithmetic

The AIR authors made a bold, imaginative decision when they opted for delinquency rates (rather than simple post-release failure) as a measure of program impact. They may have let themselves in for more trouble than they realized, however. The social science community is not in total agreement as to which techniques should be used for rate analyses, so from the start, it is impossible to please all of the people all of the time. Yet one of the analytical problems we have seen in the AIR report raises questions at a lower, much less philosophical level. The arithmetic definition of rates used in the AIR report interacts with the threat of mortality so as to bias the estimate of suppression.

To illustrate how this happens, consider the definition of delinquency rate for an individual delinquent. This is

$$\text{Rate} = \frac{\text{Number of Police Contacts in a Period of Time}}{\text{Length of Time Period}}$$

So if Time is measured in months, this rate is "police contacts per month." A post-release reduction in this rate is the basic unit of suppression.

While this definition of delinquency rate seems simple enough, it requires at least one important but arbitrary decision. What length of time should be used in the denominator of the rate? Should we start counting Police Contacts at the month of birth? At the age of six? Or eight? Or from the month of the first police contact? Of course, for pre-intervention rate estimates, and thus for a measure of suppression, these definitions of Time will give different answers.

As the AIR authors deal only in "phase" time, not in real time, they are not forced to confront this issue. In "phase" time, everything is relative to the month of intervention. To estimate a pre-intervention rate for a youth, the AIR authors count backwards one, two, or four years from the month of intervention. Their "phase" time convention assumes (incorrectly) that the time of intervention and the lengths of pre-intervention career are roughly the same for all of these delinquents. Of the data we have analyzed, however, the pre-intervention careers range in length from one to 102 months. There is so much variance along this dimension (and along others) that the activity of only a few delinquents overwhelms the estimate of group rate.

But the larger issue concerns the relationship of an individual's rate to the rate of his group. The problem is that there are at least two ways to estimate the group rate. First, the group rate can be estimated as the average of its member's rates. That is

$$\text{Group Rate} = \frac{\text{Rate}_1 + \text{Rate}_2 + \ldots + \text{Rate}_N}{N}$$

for a group of N delinquents. Second, the group rate can be estimated on the basis of the "time at risk" of its members. That is

$$\text{Group Rate} = \frac{\text{Total Number of Offenses in a Year}}{\text{Person-Months in that Year}}$$

A person-month is the number of months each person has spent at risk in the year of the estimate. For example, if one delinquent has been free for six months and another delinquent has been free for nine months, their total time at risk is (six months + nine months) 15 person-months. In the AIR report, group rates are always estimated in this second way.

These two methods of estimation yield roughly the same estimates of group rate when all the times at risk are roughly equivalent. The AIR data do not have equivalent follow-up times, but vary from 0 to over 40 months.

To illustrate the shortcoming of the estimation method, we refer to Figure 8.12, page 147 of the AIR report. The annual rate of police contact for the DOC group in the first post-release year is given here as 1.6. The AIR authors have used the second method for this estimate. Using instead the first method, we estimate this rate as 1.825 police contacts per year. And if we use the entire post-release period of time for this estimate (rather than only the first post-release year) the estimate is 2.0 police contacts per year. Our estimate of post-release delinquency changes by 25% when we use the first method.

In every case, the first method leads to a higher estimate of post-release delinquency than the second method. The reason for this discrepancy is that the second method interacts with mortality. When person-months are used in the denominator of the rate, we have to assume that two delinquents who are free for six and nine months respectively (15 person-months) will have the same average rates as two delinquents who are free for one and 14 months respectively (15 person-months). This is not generally true, of course. Delinquents who are free for longer periods of post-release time tend to be the least serious or least active delinquents.

What is the magnitude of this bias? In Chapter 9 of the AIR report, UDIS First Placements are broken down into Levels I, II, and III and then compared for the first 12 post-release months. However, only 34 UDIS delinquents have been free for 12 post-release months. The remainder of the sample (157 delinquents) had been free for less than 12 months by December 31, 1976. In Figure 9.2 (and subsequent Figures of Chapter 9), post-release rates of delinquency are estimated on the basis of 87.8 person-years or 1054 person months. By number and length of follow-up, these person-months are distributed as

Length of Follow-up	Person-Months at Risk in the Rate
34 @ 12 months	408 person-months or 45.6% of the total
9 @ 11 months	99 person-months or 9.4% of the total
2 @ 10 months	20 person-months or 1.9% of the total
9 @ 9 months	81 person-months or 7.7% of the total
8 @ 8 months	64 person-months or 6.1% of the total
8 @ 7 months	56 person-months or 5.3% of the total
10 @ 6 months	60 person-months or 5.7% of the total
3 @ 5 months	15 person-months or 1.4% of the total
13 @ 4 months	52 person-months or 4.9% of the total
10 @ 3 months	30 person-months or 2.8% of the total
9 @ 2 months	18 person-months or 1.7% of the total
10 @ 1 month	10 person-months or 1.0% of the total
66 @ Less than one month	140 person-months or 13.3% of the total

The "rounding error" has been assigned entirely to the last group: 66 delinquents who have been free for less than one month cannot possibly have 140 person-months at risk. This convention is conservative.

There is only one conclusion possible from an examination of this distribution of person-months. The post-release rate estimate is biased in favor of the least delinquent youths of the sample and this bias is substantial. If we consider only the 34 "best" and 66 "worst" delinquents, the magnitude of bias is striking. The 34 "best" are only 17.8% of the UDIS sample and yet account for 45.6% of the person-months used to estimate the post-release rate of delinquency. The 66 "worst" are 34.6% of the UDIS sample and yet account for only 13.3% of the person-months used to estimate the post-release rate of delinquency. Due to our "rounding error" convention, these 66 "worse" actually account for much less than 13.3% of the person-months.

2.1.2 Mortality: Our Conclusion

As we have demonstrated, a substantial portion of the suppression effect reported by AIR is due to nothing more than simple mortality. At the start of the post-release follow-up period, 103 delinquents are "missing" and these 103 are the most active or most serious of all. Then throughout the post-release period, delinquents "disappear" and the most active or most serious are the most likely to "disappear." Due to the seriousness of their pre-intervention careers, some delinquents are held in programs or institutions longer than the average, and thus, are available for shorter periods of follow-up. Once released, these delinquents are more likely to commit new offenses, and thus, to "disappear" back into programs and institutions. The post-release estimates of delinquency are therefore based largely on the most "casual" delinquents of the combined samples. Finally, the mortality artifact is amplified by the arithmetic used in the AIR analyses.

As we have no data on the 103 "missing" delinquents, we cannot precisely analyze the effects of mortality on the major AIR finding. We can only say that, as we have demonstrated, the mortality artifact is substantial.

In many respects, the "disappearing" delinquents contribute more to the mortality artifact than do the "missing" delinquents. A major fault of the AIR report is that the follow-up period was closed prematurely. While the AIR report typically couches its major finding in terms of "annual rates of police contact," only 141 delinquents out of 487 were free for an entire year. This is only 28.9% of the combined samples. More important, these 141 delinquents were mostly from the Pre-UDIS Baseline sample. While the AIR report was ostensibly "about" UDIS, only 34 UDIS delinquents were followed-up for a full year.

If we had pre-intervention data for the "missing" delinquents, we could make some rather simple assumptions about these delinquents (and about the "disappearing" delinquents) and use these assumptions to "correct" for the effects of mortality. This is not as easily done with the "disappearing" delinquents alone, however. Given their follow-up distributions and their relatively small number, no "correction" is possible.

To finally solve the mortality issue, we would have to collect more data on the "disappearing" delinquents. In effect, we would have to follow-up these youths to see how many police contacts they had accumulated. This type of data collection is not economically feasible for our reanalysis. However, it is our understanding that AIR has been funded for a larger study of delinquency in Chicago. We hope that AIR will address the mortality issue more closely in this larger study. To control the threat of mortality a large sample with a long follow-up is required.

2.2 Maturation

Maturation refers generally to an effect due to nothing more than a change in the organism, a change in the physiological-psychological-sociological make-up of the delinquent, that is. If we were to measure the heights and weights of these delinquents before and after the intervention, for example, we would conclude that both the UDIS and non-UDIS treatments "caused" these delinquents to grow.

There is a substantial body of literature which suggests that delinquent behavior is a function of age, that delinquent behavior becomes less prominent as the youth matures. Many of these delinquents aged a year or more in the time between the incident offense and release from programs or institutions, and during this year, we assume that maturation had some effect. In Figure 10.4, page 170, the AIR authors have broken down the suppression effect into three age categories: 14 years old or less, 15 years old, and 16 years old or more. From this analysis, the AIR authors note on page 169:

> "Evidence of some maturation effect can be inferred. The overall suppression effect during the first year after release became progressively larger as the boys got older. Similarly, pre-intervention rates of activity dropped as age rose. But the strength of the maturation effect appeared to be small relative to the suppression effect."

But <u>how</u> small?

This remark by the AIR authors must be contrasted with the statement of Dr. Charles A. Murray to the U.S. Senate Subcommittee on Juvenile Delinquency. To the Subcommittee, Dr. Murray seems to imply unequivocally and emphatically that the suppression effect is not biased by maturation:

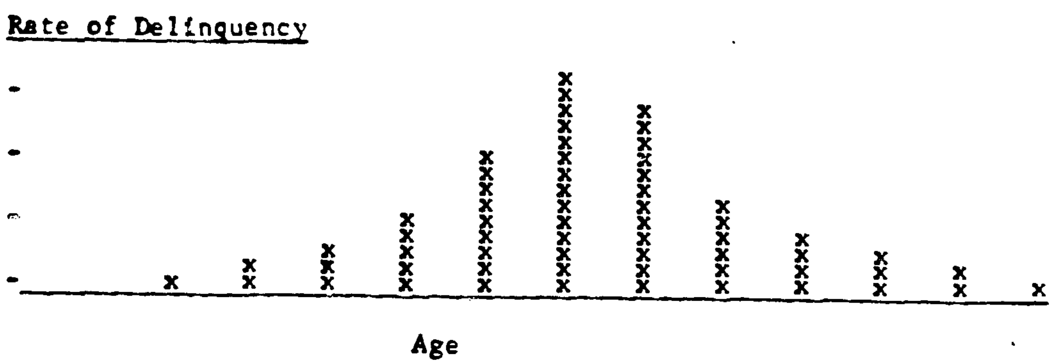
"These reductions (in rates of delinquency) were not the result of maturation; they were not the result of a delinquent career having run its course..."

and so forth. This statement is contradicted by the AIR report itself. But more important, the analysis of maturation summarized in Table 10.4 is inappropriate. A stronger, more appropriate analysis shows that the maturation effect is not as "small" as the AIR report would have its audience believe.

The basic problem with the AIR analysis of maturation is that maturation is a function of age and age is a continuous variable. By categorizing age, the AIR analysis models the maturation effect only approximately. The decision to have exactly three categories of age was no doubt dictated by statistical concerns such as sample size. Four age categories would have been better than three, five age categories would have been better than four. As the number of age categories increases, however, the sample size in each category decreases.

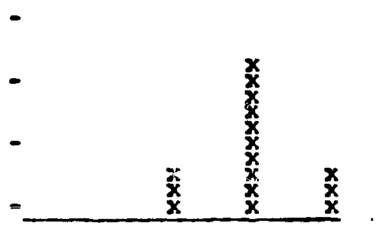
Likewise, the boundaries of the three categories are arbitrary. What is unique about the 180th month of life? Yet it is at this month, or so the scheme of categories implies, that the delinquent career begins to change profoundly. In the 180th month, the start of the 16th year, maturation begins in earnest.

To illustrate the conceptual problem with this analysis, consider the hypothetical maturation function diagrammed as



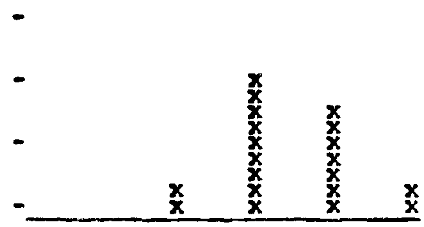
We see that the rate of delinquency increases with age, up to a point, and then begins to decrease. But when these twelve age-points are categorized (and means computed for each category), the maturation function is distorted.

Mean Rate of Delinquency



Three Age Categories

Mean Rate of Delinquency



Four Age Categories

This illustration demonstrates the problem. As the number of categories increases (up to twelve in this illustration), the scheme of categories comes closer and closer to the actual maturation effect. However, with a smaller number of categories, say three, the distortion is substantial and weakens the analysis.

As a related problem, categorization requires some knowledge of the maturation function. Each category should capture some point of the function's inflection. That any maturation effect at all was discovered by the AIR analysis suggests that the maturation effect is quite large.

2.2.1 A Reanalysis of the Maturation Effect

In light of our comments on the AIR analysis of maturation (as summarized in their Table 10.4), it is clear that age must be handled as a continuous variable. For analytical purposes, we hypothesize that the post-release rate of delinquency for an individual is a function of the individual's age at the time of release from program or institution. This can be expressed as

$$\text{Suppression} = b_0 + b_1(\text{Age at Release})$$

Other things equal, we expect older delinquents to have lower post-release rates of police contact than younger delinquents. The parameter b_1 is thus expected to be negative.

To conduct our reanalysis, we have defined Suppression as

$$\text{Suppression} = \frac{\text{Total Arrests in First Post-Release Year}}{\text{Total Arrests in Last Pre-Intervention Year}}$$

If a delinquent's post-release rate of arrest is lower than his pre-intervention rate of arrest, then

$$0 \leq \text{Suppression} < 1$$

that is, Suppression will be a fraction. If a delinquent's post-release rate of arrest is exactly the same as his pre-intervention rate of arrest, then

$$\text{Suppression} = 1$$

which implies that the intervention had no effect. Finally, if the delinquent's post-release rate of arrest is higher than his pre-intervention rate of arrest,

$$\text{Suppression} > 1$$

When Suppression is greater than unity, the implication is that the intervention has been harmful; it has made the delinquent more delinquent. When Suppression is less than unity, when it is a fraction, that is, the implication is that the intervention has been helpful; it has made the delinquent less delinquent.

In the AIR report, the suppression effect is ordinarily stated in terms of "percent reductions" in arrest rates from pre-intervention to post-release. The relationship between Percent Reduction and Suppression is simply

$$\text{Percent Reduction} = 100\% - 100 \times \text{Suppression}$$

$$\text{Suppression} = 1/100 \times (100\% - \text{Percent Reduction})$$

So we are operating in the same metric as the AIR report. For analytical purposes, however, it is easier to operate with Suppression than with Percent Reduction.

Because Suppression is constrained to the interval

$$0 \leq \text{Suppression} < +\infty$$

we must transform Suppression. The appropriate transformation is the natural logarithm transformation. The natural logarithm (denoted by "Ln") of Suppression is constrained to the interval

$$-\infty < \text{Ln(Suppression)} < +\infty$$

Because Ln(Suppression) may take on any value between negative and positive infinity, we are free to use regression methods on it. But our model is now

$$\text{Suppression} = b_0(\text{Age at Release})^{b_1}$$

$$\text{Ln(Suppression)} = \text{Ln}(b_0) + b_1\text{Ln(Age at Release)}$$

We will report our results in both the standard and natural log metrics.

Of the 487 delinquents analyzed by AIR, only 141 were followed-up for at least one year. The average or mean value of Suppression for these 141 delinquents is

$$\text{Suppression} = .3923$$

or $$\text{Percent Reduction} = 60.77\%$$

which is a substantial Percent Reduction.

We must now consider how the Suppression scores are distributed across the delinquent population by Age at Release. The estimated regression parameters for our model are

$$\text{Ln(Suppression)} = \underset{(4.1826)}{6.9243} - \underset{(.7920)}{1.4885}\ \text{Ln(Age at Release)}$$

with standard errors (4.1826) (.7920)

For these 132 delinquents, the average Age at Release is 196.51 months or 16.38 years old. Substituting this mean into the model gives us

$$\begin{aligned}
\text{Ln(Suppression)} &= 6.9243 - 1.4885\ \text{Ln(Age at Release)} \\
\text{Ln}(.3923) &= 6.9243 - 1.4885(196.51) \\
-.9357 &= 6.9243 - 1.4855(5.2807) \\
-.9357 &= 6.9243 - 7.8601
\end{aligned}$$

We may interpret these numbers to mean that the effect of maturation is greater than the effects of all other variables combined.

For the moment, consider a delinquent who is one year younger than the average at the time of release. For this delinquent, the expected Ln(Suppression) is

$$\text{Ln(Suppression)} = 6.9243 - 1.4885\ \text{Ln(184.51 Months)}$$
$$= 6.9243 - 1.4885(5.2177) = -.8266$$

which gives us

$$\text{Suppression} = e^{-.8266} = .4375$$
$$\text{Percent Reduction} = 56.25\%$$

This is still a substantial Percent Reduction but a smaller effect than average. As Age at Release becomes lower, Suppression and Percent Reduction are expected to become smaller and smaller. Finally, a delinquent who is only 104.78 months old at release has an expected

$$\text{Ln(Suppression)} = 6.9243 - 1.4885\ \text{Ln(104.78 Months)}$$
$$= 6.9243 - 1.4885(4.6519) = 0$$

which gives us

$$\text{Suppression} = e^{0} = 1$$
$$\text{Percent Reduction} = 0\%$$

Delinquents younger than 104.78 months or 8.73 years at release are expected to have a negative Percent Reduction, that is, are expected to have higher arrest rates post-release than pre-intervention. As there were no delinquents in the sample who were this young at time of release, there were no delinquents who were expected to have negative Percent Reductions --- although a number of delinquents did.

Now consider a delinquent who was older than the average at the time of release. A delinquent who was one year older than the average would have an expected

$$\text{Ln(Suppression)} = 6.9243 - 1.4885\ \text{Ln(204.51 Months)}$$
$$= 6.9243 - 1.4885(5.3206) = -.9954$$

which gives us

$$\text{Suppression} = e^{-.9954} = .3696$$

Percent Reduction = 63.04%

As a delinquent's Age at Release increases then, the expected Percent Reduction becomes larger. For ages at six-month increments from 180 months, the expected Percent Reductions are

180 months or 15 years:	Percent Reduction	= 55.31%
186 months	Percent Reduction	= 58.11%
192 months or 16 years:	Percent Reduction	= 59.40%
198 months	Percent Reduction	= 61.22%
204 months or 17 years:	Percent Reduction	= 62.91%
210 months	Percent Reduction	= 64.47%
216 months or 18 years:	Percent Reduction	= 65.93%
222 months	Percent Reduction	= 67.29%
228 months or 19 years:	Percent Reduction	= 68.57%

The oldest Age at Release in our sample of 141 delinquents is 228 months, or 19 years old. Overall, the ages are skewed upwards. The youngest Age at Release is 149 months, or 12.42 years old. The modal Age at Release is 208 months, or 17.33 years old. The median Age at Release is 198 months or 16.5 years old.

The strong and regular relationship between Age at Release and Suppression can be most parsimoniously explained as a maturation effect. While all 141 delinquents in our sample are expected to have substantial Percent Reductions, the larger Percent Reductions belong to the oldest delinquents in the sample. The substantial size of the Percent Reduction of the entire sample is due largely to the age structure of the sample. In the next section, we will add another independent variable to our model to account for a regression artifact. When both maturation and regression are controlled statistically, we arrive at a somewhat more realistic picture of the "suppression effect" reported by AIR for these delinquents.

2.3 Regression

Regression as a statistical threat to internal validity occurs when delinquents have been selected for a correctional intervention. To the extent that their selection is due to a high pre-intervention rate of delinquency, some post-release reduction can be expected which is not attributable to the intervention. To demonstrate regression, consider the roll of two honest dice. On a given roll, the possible numbers by combination are

2: (1,1)
3: (1,2) (2,1)
4: (1,3) (2,2) (3,1)
5: (1,4) (2,3) (3,2) (4,1)
6: (1,5) (2,4) (3,3) (4,2) (5,1)
7: (1,6) (2,5) (3,4) (4,3) (5,2) (6,1)
8: (2,6) (3,5) (4,4) (5,3) (6,2)
9: (3.6) (4.5) (5,4) (6,3)
10: (4,6) (5,5) (6,4)
11: (5,6) (6,5)
12: (6,6)

considering this probability space, it is obvious that the number seven will come up more often than any other number; the numbers two and twelve will come up the least often of any numbers.

Now suppose that we roll many pairs of dice and select for "treatment" only those pairs which have come up on a first roll with a high number: any number larger than seven. Let us then "treat" these high-rolling dice by uttering a magic phrase over them. A second roll is likely to produce a smaller number, and, in effect, we have "cured" the dice. We have not "caused" a true suppression effect, however, but rather, have fallen prey to a regression artifact.

The magical "treatment" works only because the number seven is the "expected value" on a roll of dice. If a larger number comes up on a first roll, a smaller number on the second roll is the likely expectation. The probabilities are as follows

First Roll	Probability of a Smaller Number on the Second Roll
8	21/36 = .5833
9	26/36 = .7222
10	30/36 = .8333
11	33/36 = .9166
12	35/36 = .9722

The higher the number on the first roll, the greater the probability of a reduction on the second roll. We will use this principle in our analysis of the contribution of the regression artifact to the suppression effect.

The AIR authors are aware of the regression artifact in their data. On page 164 they note:

> "In the case of correctional interventions for delinquents, some regression effect must be assumed; the fact that an intervention took place implies either an unusually high or a serious level of delinquent activity at that time. It is a classic setting for a regression effect. But we question whether the proportion is large."

But as in the case of maturation, the AIR report describes a inadequate model for evaluating or measuring the effects of regression. In Figure 10.1, page 163, the AIR authors fit time series regression lines to the pre-intervention and post-release rates of delinquency. On page 163, they explain:

> "The difference between the levels of pre and postintervention activity are striking...We label the phenomenon the "suppression effect..."

They then note the possibility of a regression artifact. To rule this possibility out, they delete the last six pre-intervention observations and estimate the suppression effect as shown in Table 10.1, page 165. As the suppression effect does not vanish entirely, they conclude that the effect of regression is small. Our reanalysis will disagree with this conclusion.

It is important to remember in interpreting the AIR report that they have used suppression rates to index recidivism and that this is an uncommon measure of program success. Using simple success/failure, where any subsequent re-arrest betokens complete failure, the AIR findings are consonant with the many other studies using that more conventional measure of recidivism: most of these youth are re-arrested in the post-release period, and the result merely supports the conventional wisdom of delinquency research: NOTHING WORKS (Martinson, 1976; Lipton, Martinson & Wilks, 1975; Schur, 1973).

AIR's substitution of comparative rates of delinquency (pre- vs post- intervention) for simple post-incarceration failure appears judicious since simple failure does not distinguish those whose delinquency has extinguished from those whose activity has merely diminished. A rate measure captures such distinctions and seems particularly justified for youths from neighborhoods where police contact is almost commonplace. To reduce delinquency substantially, even if not to extinction, is worthy of note, and a comparison of rates considers the entire pattern of post-intervention behavior.

AIR claims that it is this increased sensitivity of their index which accounts for their finding in apparent contradiction to the bulk of delinquency intervention research. And while most research concludes that intervention is a failure, previous studies using rates of inhibition to gauge program impact have reported substantial reductions in post-release delinquent behavior. In both Empey & Lubeck (1971) and Empey & Erickson (1972), pre-intervention rates of police contact increase almost exponentially, and then drop abruptly and profoundly in the post-release period, corresponding to the AIR data in Figure 1.1 (page 7 of this report). Given the gross differences among these many programs, a new conventional wisdom is emerging: when supppression rates are used to measure program impact, ANYTHING WORKS.

Since AIR argues that no artifacts are implicated in their finding, they use "suppression effect" to refer to the entire drop in police contacts, and not just to that portion of the drop which is due to intervention alone. We show that a compelling argument for intervention cannot be made from their analysis, or from these data.

The AIR authors have noted that the pre-intervention arrest rates for these delinquents trend exponentially upwards to the time of intervention. They have used this appearance of the data to argue that no regression artifact is operating. Ironically, it is precisely this facet of the data which indicates a regression artifact in operation. Maltz and Pollock (1978, appended) have shown that when juvenile court judges select delinquents for intervention, a pre-intervention trend of this sort is to be expected.

In the next sections, we describe arrest rates over time as a random process. The rates fluctuate randomly about some mean level, sometimes higher and sometimes lower than the mean. Delinquents are not selected for intervention as a result of an abnormally low arrest rate, but rather, as a result of an abnormally high rate. This selection criterion guarantees an exponentially increasing arrest rate in the pre-intervention period as well as a sharp drop in arrest rates during the post-release period. After demonstrating this phenomenon, we reanalyze the AIR data, concluding that in these data the apparent suppression effect could have been achieved from the impact of regression and maturation alone. Their analysis does not demonstrate that the intervention *per se* was effective.

2.3.1 Police Contacts as a Random Process

The AIR authors call their sample of youths "chronic delinquents." This means operationally that each youth is arrested with some regularity, and thus, that each has a non-zero rate of arrest over time. Given the many variables that determine whether an officially recorded arrest will occur, it is reasonable to assume that this rate is random. We are not arguing here that delinquency per se is a random process, that these particular youths are the unfortunate losers in a sociological game of chance. Rather, accepting for this analysis that these youths are indeed delinquents, and given the nature of delinquency and the nature of arrests, we are arguing that the timing of each arrest is random.

With this assumption, police contacts in time can be well described as a Poisson process. The probability that a delinquent will have exactly k arrests in an interval of time t is then given by

$$P(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

The Poisson parameter, λ, is the expected rate of arrest during the interval t. A maximum likelihood estimate of λ is given by

$$\lambda = \frac{\text{Number of Arrests}}{\text{Time}}$$

where the unit of time is arbitrarily chosen so that $0 < \lambda < 1$.

Using $\lambda = .33$ for these delinquents would imply, for example, that each expects to be arrested four times in a year, and that the probability of <u>exactly</u> <u>k</u> arrests in any given month is

$$
\begin{aligned}
P(0) &= .71653 \\
P(1) &= .23884 \\
P(2) &= .03981 \\
P(3) &= .00423
\end{aligned}
$$

and so forth.

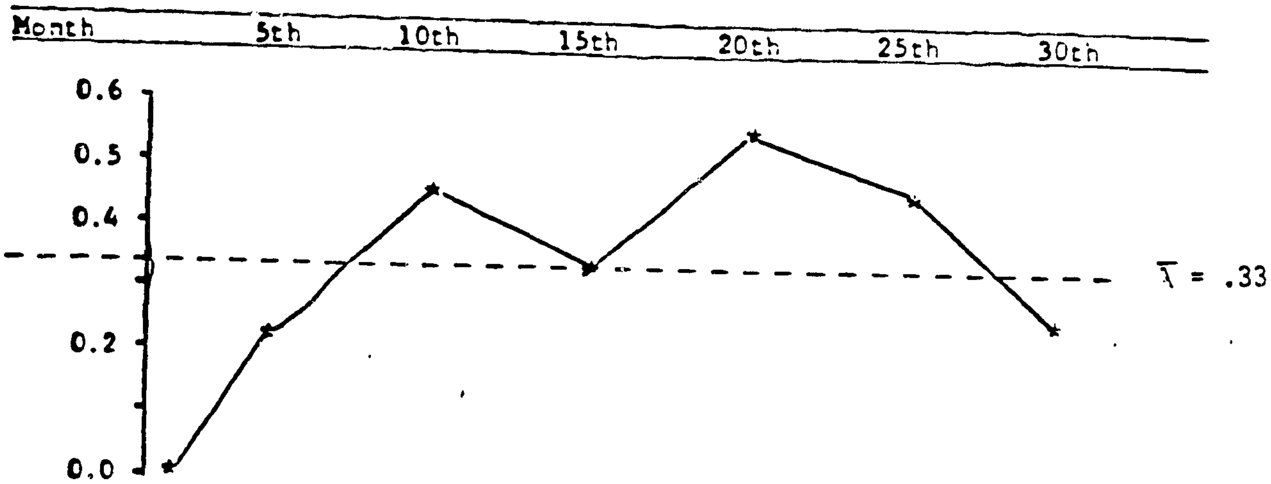
One interpretation of these probabilities is that an <u>average</u> delinquent career will be composed of independent months. Approximately 71.7% of these months will have <u>no</u> arrests; 23.9% will have <u>one</u> arrest; and slightly less 4% will have <u>two</u> arrests. To demonstrate what this means, consider a thirty-month long delinquent career generated randomly as shown below. Using $\lambda = .33$, we expect ten arrests during the career. The following would be one typical distribution:

```
                                x                                   x
                                                                    x
 . . . . . x x . . . . x . . . . . . x . x . . . x . . x . . .
___________________________________________________________________
Month #      5         10         15         20         25         30
```

Using the entire thirty-month career, we see that

$$\lambda = \frac{10 \text{ Arrests}}{30 \text{ Months}} = .33$$

Yet for shorter periods of time, λ <u>fluctuates</u> <u>wildly</u> about this mean. Using a nine-month moving average to estimate λ , we see that the rate parameter changes with time:



During some brief time spans, the youth would appear highly delinquent; during other, hardly delinquent at all. But this is merely a function of chance, and of the time frame used to assess the delinquent career.

We may now ask <u>how</u> juvenile court judges select delinquents for intervention. Few of the delinquents studied by AIR were first-offenders. Most had extensive police records prior to selection for intervention. What was so different about the precipitating offense? One might anticipate that the precipitating offense was somehow "more serious" than prior offenses, but on the average, the precipitating offense for these delinquents was no more serious.

We may surmise nevertheless that judges use some selection <u>criterion</u> and that this criterion involves an attempt to estimate the delinquent career objectively. Using one plausible selection criterion, the judge might tolerate a certain rate of arrest for the delinquent over some finite interval of time. But if the rate of arrest were to exceed (<u>the straw that broke the camel's back</u>) the limit of toleration, the judge would sentence the delinquent to an intervention.

This strategy can be operationalized as the criterion of <u>k</u> arrests in the interval $\underline{t}$. For example, if a delinquent appears before the judge two ($\underline{k}$ = 2) times in six (t = 6) months, the judge decides to sentence the delinquent to intervention. The second offense is the straw that broke the camel's back. Note also that when the judge uses this strategy, he is estimating a value of λ and his estimate is based on a particular time frame.

This hints at what will occur during the postintervention period. For an expected arrest rate of λ = .33 the probabilities of exactly k arrests are in a 6 month interval are

$$
\begin{aligned}
P(0) &= .135\\
P(1) &= .271\\
P(2) &= .271\\
P(3) &= .180\\
P(4) &= .092\\
P(5) &= .036\\
P(6) &= .012
\end{aligned}
$$

and so forth.

We interpret these probabilities to mean that 13.5% of the delinquents in this cohort will have no arrests in a given six-month period; 27.1% will have exactly one arrest; 27.1% will have exactly two arrests; and 18% will have exactly three arrests. This variance occurs by chance alone yet it would appear that some of the delinquents are "better" or "worse" than others.

We can use these probabilities to determine the level of police contact in the six-month period following intervention. When this expected level is higher than the preintervention level, then the intervention appears to have "caused" the delinquents to become more delinquent. When the expected level is the same, then the intervention appears to have had no effect. And if the expected level is lower, then the intervention appears to have "suppressed" the delinquent behavior. If the correctional intervention *per se* had no effect on delinquency rates, the probabilities are

6 MONTHS PREINTERVENTION		6 MONTHS POST INTERVENTION if $\lambda = .33$		
No. of Police Contacts	Probability of this No.	Probability of Higher No.	Probability of Same No.	Probability of Lower No.
k = 0:	P(0) = .135	.865	.135	.000
k = 1:	P(1) = .271	.594	.271	.135
k = 2:	P(2) = .271	.323	.271	.406
k = 3:	P(3) = .180	.143	.180	.677
k = 4:	P(4) = .092	.051	.092	.857
k = 5:	P(5) = .036	.015	.036	.949
k = 6:	P(6) = .012	.003	.012	.985

This is the same regression phenomenon we outlined in the introduction to this section. There we used a dice-rolling example, showing that the higher the number on the first roll, the more likely a reduction in that number on a second roll.

The same principle holds when we consider delinquency rather than dice-rolling. Given a treatment that has no effect whatsoever, selection for high rates of pre-intervention activity is likely to make the treatment look effective. Moreover, the more selective (or lenient) the judge, the greater the regression artifact. Judges who wait until the third offense, for example, select only 18.8% of the delinquents who are available for selection. The probability that these delinquents will have a lower rate of arrest after release is .677. On the other hand, judges who wait until the fourth offense will select only 14.3% of the delinquents and the probability that these delinquents will have a lower post-release rate is .857.

We will use this principle in our reanalysis.

A Simulation

In Figure 2.3.1(2) on the following page, we show the monthly arrest rates for a combined sample of N = 365 delinquents from AIR's data. In Figure 2.3.1(3) on page 43, we show simulated data, generated in the following manner:

(1) Using a Poisson rate parameter of $\lambda = .33$, 96 digits were generated, each representing the number of arrests in a single month. This series of random digits is taken as the delinquent career. In all, 1000 delinquent careers were generated.

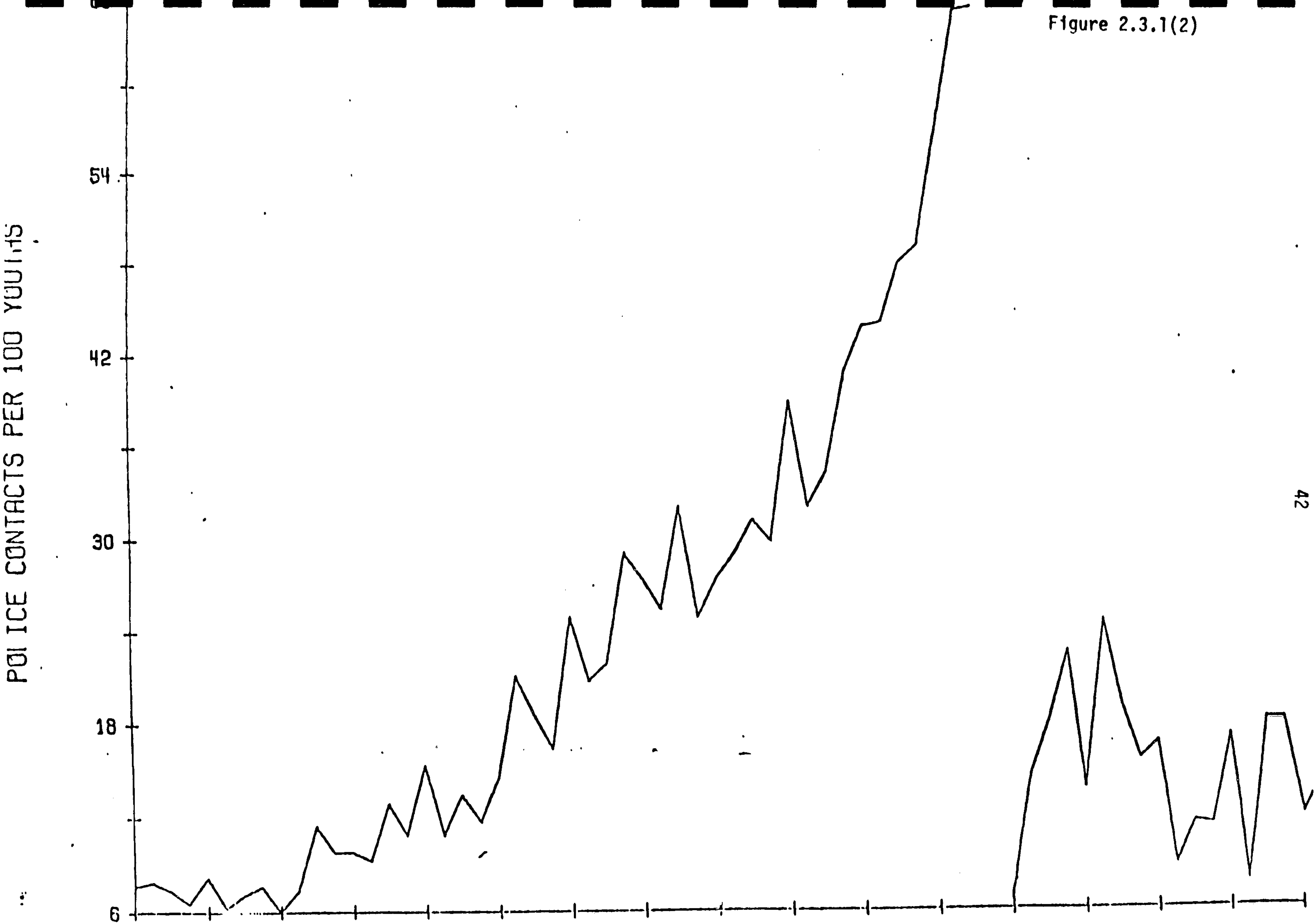
(2) Whenever a delinquent has one or more arrests in a month, he is taken before a judge. If the incident arrest is sufficiently "serious," the judge sentences the delinquent to a correctional intervention. Two arrests out of each 100 were deemed this serious.

(3) For all other arrests, the judge bases his sentencing decision on an evaluation of the delinquent's immediately prior record. The length of time defined as "immediately prior" was random, distributed uniformly from two to 20 months. The judge counted the number of arrests during this period and computed an arrest rate. If the rate was higher than a random criterion (.33 +, distributed uniformly), the delinquent was sentenced to a correctional intervention. If not, the delinquent was returned to the community until his next arrest.

(4) After all 1000 delinquents had been selected, the data were aggregated and plotted as shown in Figure 2.3.1(3) on page 43.

These data, simulated under the assumption that correctional intervention has no independent effect, show a regression artifact which makes the correctional intervention appear effective. The effect is strikingly similar to that seen in the actual data. That some "suppression effect" remains in

Figure 2.3.1(2)



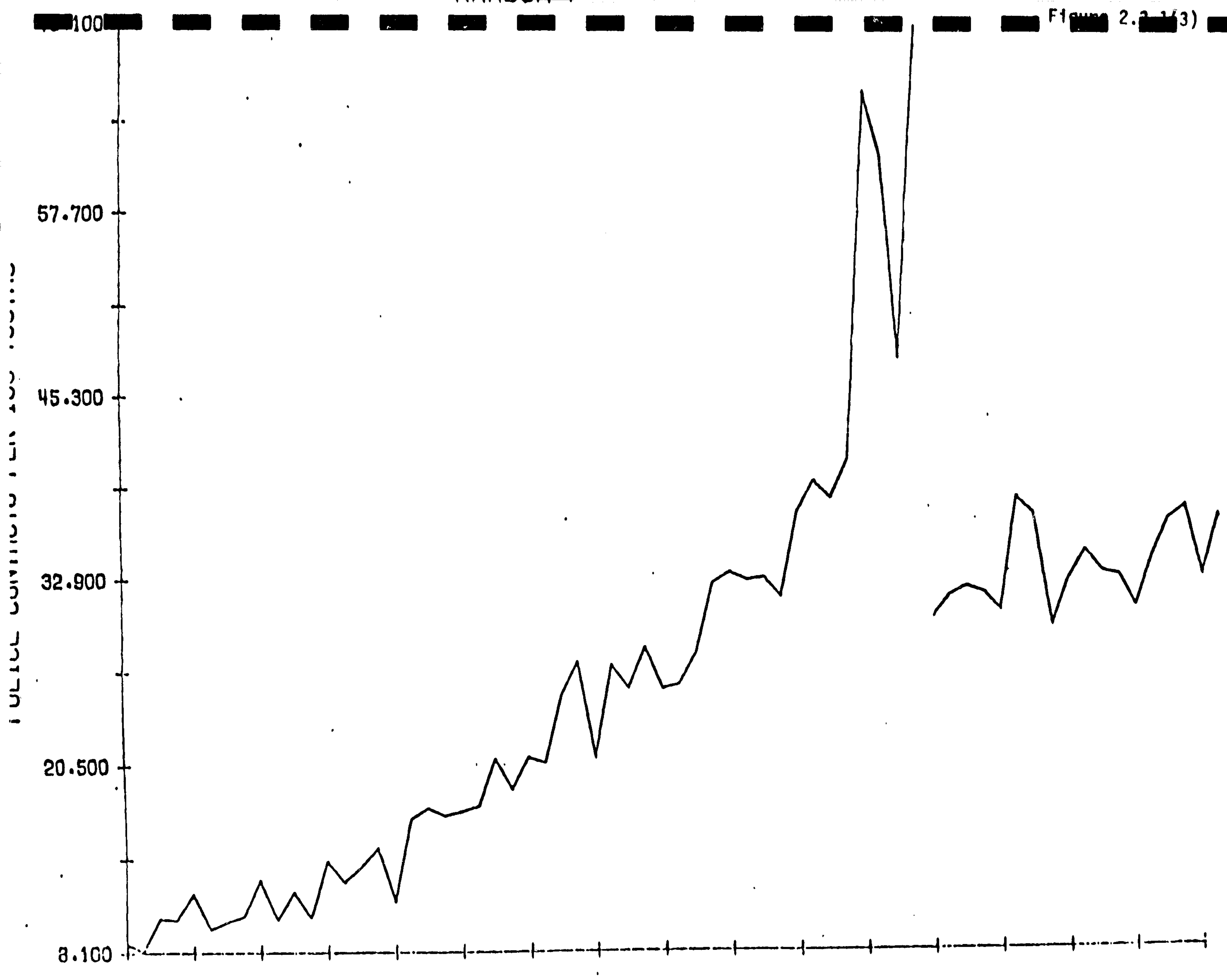


Figure 2.3.1(3)

the real data can be explained by two other factors. First, in the real data maturation decreases the delinquency rate. The delinquents were kept in institutions and programs for nearly a year on the average. As we have shown, maturation out of delinquency was substantial during this period. Second, the real data are affected by mortality. Successive post-release delinquency rates are computed on the basis of fewer and fewer delinquents. This makes the post-release rates unstable. (When these two factors are held constant, the simulated data and the real data show much the same effect.)

2.3.2 Our Reanalysis of the AIR Data

We have defined Suppression as the ratio of arrests in the first post-intervention year to arrests in the last pre-intervention year. That is

$$\text{Suppression} = \frac{\text{Total Arrests in First Post- Year}}{\text{Total Arrests in Last Pre- Year}}$$

For a given delinquent then, Suppression will always be a positive number. If Suppression is a fraction, the delinquent has been arrested less frequently in the first post-intervention year than in the last pre-intervention year.* If Suppression = 1, pre- and post-intervention arrests are equal, and for that delinquent, the intervention has had no impact. Finally, if Suppression is greater than unity, the delinquent has had more post-intervention arrests than pre-intervention arrests. In the AIR report, effects are generally stated as Percent Reductions. This measure of recidivism is related to Suppression by

$$\text{Percent Reduction} = 100 - 100 \times \text{Suppression}$$

For analytic purposes, we will deal with Suppression scores, but of course, our findings pertain as well to Percent Reduction scores.

To understand the relationship between a regression artifact and Suppression or Percent Reduction, we must understand the relationship between the Percent Reduction of an individual delinquent and the Percent Reduction of the cohort. If a cohort experiences a 50% reduction in arrests from pre- to post-intervention, and if there is no regression artifact, this means that, on the average, every delinquent in the cohort has experienced the same Percent Reduction: 50%. In particular, if we divide the cohort into halves, one half with higher than average pre-intervention arrest

*We use only the last pre-intervention year, as do the AIR authors, because this statistic is nonstationary. The longer the pre-intervention segment used in this measure, the lower the Suppression score.

rates and the other with lower than average pre-intervention arrest rates, both halves of the cohort are expected to have the same Percent Reductions.

On the other hand, if there is a regression artifact in operation, we will see a strong relationship between pre-intervention arrests and Percent Reductions. Delinquents with higher than average pre-intervention arrest rates will experience higher than average Percent Reductions. Delinquents with lower than average pre-intervention arrest rates will experience lower than average Percent Reductions. To rule out a regression artifact in these data then, we need only examine the covariance structure between these two variables.

As we have already noted, the AIR findings are biased in favor of a large Suppression effect by the threat of mortality. In particular, delinquents who commit serious crimes during the post-intervention period are reincarcerated. Mortality also presents a problem for our reanalysis. Because the AIR sample of delinquents has been ravaged by mortality, and because mortality has weeded the most active delinquents, we must examine subsamples to get an accurate picture of the regression artifact.

(1) The first subsample we will examine is the N = 91 delinquents who were followed up by AIR for at least one postintervention year and who had at least one police contact during that time. This excludes 50 delinquents who were followed up by AIR for at least one postintervention year but who had no police contacts. We will analyze these delinquents in the next subsample.

As the dependent variable of our reanalysis, Suppression, ranges from zero to some small positive number, we must transform the Suppression scores into the natural logarithm metric. Our multiple regression model is thus

$$\text{Suppression} = e^{(b_0 + b_1x_1 + b_2x_2 + b_3x_3)}$$

Our first independent variable, x_1, is the square root of pre-intervention arrests. The number of pre-intervention arrests for these delinquents has a negatively skewed distribution, so the square root transformation makes the distribution more normal. This is a standard multiple regression technique.* <u>If there is no regression artifact operating in this subsample, we expect the parameter b_1 to be zero</u>.

Our second independent variable, x_2, is the age in months at release from program or institution. Due to maturation, we expect the parameter b_2 to be negative. In words, this means that we expect older delinquents to experience a greater <u>Percent Reduction</u> than younger delinquents.

Finally, our third independent variable, x_3, is the number of months served in institutions or programs for the incident offense. The rationale given in the AIR report for the <u>Suppression</u> effect is specific deterrence. We thus expect the parameter b_3 to be negative. Delinquents who serve longer periods of time in programs or institutions should experience greater <u>Percent Reductions</u> because, in those cases, the deterrent agent has been more severe.

Parameter estimates for this subsample of delinquents are

Ln(Suppression)	=	3.25	-	.954 x_1	- .011 x_2	-	.015 x_3
Standard Errors		1.09		.116	.005		.014

The standard errors printed below the parameter estimates may be used to test the parameters for statistical significance. For a .05 level of significance, we require a parameter estimate to be twice the size of its standard error. We see by this criterion that the parameter b_3 is not statistically significant. All other parameters are statistically significant and explained variance is high, $R^2 = .4477$

To interpret this model, we note that the average natural logarithm of <u>Suppression</u> for these N = 91 delinquents is

*This transformation has no effect on the results we report here. The parameter estimates and conclusions are much the same with an untransformed x_1.

$$\text{Mean Ln(Suppression)} = -.8967$$

$$\text{so} \quad \text{Mean Suppression} = e^{-.8967} = .4079$$

$$\text{Mean Percent Reduction} = 59.21\%$$

We must now examine the distribution of <u>Percent Change</u> scores about this mean. We do this by substituting the mean age at release (197.3 months old) and mean time served (8.439 months) into the model.

$$\begin{aligned} \text{Ln(Suppression)} &= 3.25 - .954x_1 - .011(197.3) + .015(8.439) \\ &= 3.25 - .954x_1 - 2.17 + .127 \\ &= 1.207 - .954x_1 \end{aligned}$$

Now consider a delinquent who has only one police contact in the last pre-intervention year. The <u>Percent Reduction</u> expected for this delinquent is

$$\text{Ln(Suppression)} = 1.207 - .954\sqrt{1} = .253$$

$$\text{Suppression} = e^{.253} = 1.287$$

$$\text{Percent Reduction} = -28.79\%$$

For this delinquent, we expect a negative <u>Percent Reduction</u>. That is, we expect this delinquent to have 28.79% more post-intervention arrests than pre-intervention arrests! In fact, we can demonstrate that only those delinquents with five or more arrests in the last pre-intervention year will have above average <u>Percent Reduction</u> scores.

For this subsample of N = 91 delinquents, we can only conclude that the regression artifact is substantial. To rule out a regression artifact, we would expect our estimate of b_1 to be zero. In fact, it is not zero. The substantial <u>Percent Reduction</u> observed for these delinquents is due largely to the presence of a few delinquents with many pre-intervention police contacts. These delinquents experience much larger <u>Percent Reductions</u> than the average, and as a result, overwhelm the mean <u>Percent</u>

Reduction of the group.

But the most surprising finding for this subsample is the sign of the b_3 parameter. It is exactly the opposite of what we expected. Because b_3 is positive, we might conclude that the more time a delinquent serves in programs or in institutions, the smaller the Percent Reduction. On the average, delinquents who spent only one day in programs or institutions (none did) would experience the greatest Percent Reductions. This parameter estimate is not statistically significant, however, so we must conclude that the time served by a delinquent has no impact on that delinquent's Suppression or Percent Reduction scores. But this interpretation too is inconsistent with the AIR claim of a special deterrent effect for the UDIS and DOC treatments.

(2) We have argued elsewhere (Gordon et al, 1978) that the mortality artifact in these data is so profound that the major AIR finding must be dismissed as a matter of course. Charles A. Murray (1978) has responded by arguing that a subsample of N = 141 delinquents who were followed up for at least one year are not subject to the mortality criticism; that because a Suppression effect for these delinquents has been observed, the validity of the AIR major finding survives. We find no merit in this argument, however. Parameter estimates of our model for this subsample are

Ln(Suppression)	=	2.54	-	.117 x_1	-	.027 x_2 +	.045 x_3
Standard Errors		3.76		.396		.019	.051

For this subsample, all parameters are statistically insignificant at the .05 level and explained variance is low, R^2 = .0165. We are not surprised by this, however, and note that the signs of all parameters are the same with this subsample.

The difference between this subsample and the last subsample is that 50 delinquents who have had no police contacts in the first post-intervention year have been added.* In one sense, these 50 delinquents serve as "outliers." All 50 are clustered about the independent variable axis. In a larger sense, however, these 50 delinquents point out a fundamental weakness of the Suppression measure of recidivism.

The 50 delinquents who had no police contacts in the first post-intervention year all have Percent Reduction scores of 100%. Yet some of these delinquents had a dozen or more police contacts in the last pre-intervention year while others had only one or two contacts during the same period. A reduction of a dozen or more police contacts should be a fundamentally different impact than a reduction of only one police contact. Yet both of these absolute reductions are seen as "equal" by the Suppression measure. This presents a problem in our reanalysis because, by definition, there can be no covariance structure between pre-intervention arrests and Suppression for these 50 delinquents. To minimize this problem, we require longer follow-up periods for all delinquents. As these data are not available, we must examine more subsamples of the population.

(3) We are able to estimate annual post-intervention rates of police contact for the "disappearing" delinquents (though not for the "missing" delinquents) by extrapolating their daily rates. For all N = 365 delinquents, including those 224 who were followed up for less than one post-intervention year, we have

$$\text{Estimated Police Contacts} = 365\left(\frac{\text{Total Contacts}}{\text{Length of Follow-up in Days}}\right)$$

We will have more confidence in our estimates from the N = 141 delinquents who were followed up for at least one year, however, so we must obtain our parameter estimates through a weighted least-squares algorithm. For all N = 365 delinquents, the parameters of our model are

*As Ln(Suppression) = 0 is undefined, we have given each of these delinquents .01 post-intervention police contacts.

$$\text{Ln(Suppression)} = 2.23 - .017\,x_1 - .027\,x_2 + .008\,x_3$$

Standard Deviations	2.301	.051	.012	.029

We see that for all N = 365 delinquents, the signs and relative sizes of the parameters are the same as in our other subsamples, although only b_2 is statistically significant from zero. The explained variance for this model is also quite low, $R^2 = .0153$.

With this subsample, the fundamental weakness of these data with respect to "outliers" is magnified. Whereas in the N = 141 subsample there were 50 delinquents with no post-intervention police contacts, in this N = 365 subsample, there are 132 delinquents with no post-intervention police contacts. But in this subsample, many of the delinquents with no post-intervention police contacts were followed up for only a few days or weeks.

(4) Our final subsample consists of the N = 233 delinquents who had at least one post-intervention police contact.

Parameter estimates for this subsample are

$$\text{Ln(Suppression)} = 4.76 - .545\,x_1 - .017\,x_2 + .004\,x_3$$

Standard Errors	1.64	.173	.008	.022

For this subsample, all parameters except b_3 are statistically significant. Moreover, the explained variance of the model is remarkably high, $R^2 = .3346$.

Comparing the model results for these four subsamples gives us a good picture of what is happening to these delinquents. In all subsamples, the signs and relative sizes of the model parameters remain the same. When those delinquents who have had no post-intervention arrests are included in a subsample, the model parameters and explained variance are statistically insignificant. When a subsample excludes these delinquents, the model parameters explained variance are statistically significant. In fact, the R^2 statistics for the N = 91 and N = 233 subsamples are much higher than would be expected. In effect, we are explaining up to 45% of the variance in Percent Reductions as a function of regression and maturation artifacts.

The low R^2 statistics for the N = 141 and N = 365 subsamples are expected. As mentioned, there can be no covariance structure by definition among the model variables when Percent Reduction is 100%. We could "correct" our models for this definitional shortcoming by, for example, assuming that every delinquent would eventually experience at least one police contact. This would automatically inject variance into the Suppression scores of the delinquents who had no post-intervention police contacts, and as a result, would permit a covariance analysis. However, we are not certain that this "correction" would be fair.

2.3.3 Conclusions on Regression and Maturation Artifacts

Overall, the results of our reanalysis are straightforward. We analyzed a great many subsamples of the AIR data (not just those reported here) and each has led to the same conclusion.

(1) For those delinquents with at least one post-intervention police contact, there is a strong relationship between pre-intervention activity and Percent Reduction. The highest Percent Reduction scores belong to those delinquents who had relatively many police contacts in the last pre-intervention year. The relationship here is reinforced by the high explained variance statistics. For these delinquents, up to 45% of the variance in Percent Reduction scores can be attributed to regression and maturation artifacts.

(2) For those delinquents with no police contacts during the first post-intervention year (or less than one year for the N = 365 sample), we cannot make an estimate of these effects. By definition, there is a zero covariance structure between pre-intervention arrests and Percent Reductions. However, we suspect that the relationship would be equally strong in this sample.

(3) In every subsample we have analyzed, the time served by a delinquent in programs or institutions has a positive or zero impact on the delinquent's Percent Reduction score. We caution against any causal interpretation of this finding. However, the AIR major finding has a certain commonsense appeal based on a specific deterrence theory of cause. Yet we find that those delinquents who were "punished more severely" than average had lower Percent Reduction scores, or at least, had scores no different

than those delinquents who were "punished less severely." While subject to other interpretations, this finding casts doubt on the AIR major finding of a suppression effect. If indeed there was a suppression effect for these delinquents, we can only speculate as to the mechanism underlying the effect. We have found no evidence for a special deterrence mechanism.

In a broader sense, we have found that Suppression is a meaningful measure of recidivism in the case where every delinquent is expected to be rearrested. Suppression is less meaningful in the case where only some delinquents are expected to be rearrested. There is a fundamental difference between an absolute reduction of a dozen or more arrests from pre- to post-intervention and an absolute reduction of only one or two arrests. Yet both of these absolute reductions are "equivalent" Percent Reductions for those delinquents who are not rearrested. In the next section, we will explore a model of recidivism which posits two subpopulations of delinquents: successes and failures.

3.0 A Secondary Finding: No Difference between UDIS and DOC

A primary initial goal of the AIR evaluation was to compare the effectiveness of UDIS and DOC. As the controversial suppression effect grew in apparent significance, however, between-program contrasts faded into the background.

The major AIR "finding" (a suppression effect) was based on a broad before/after contrast, diagrammed as

0 X 0 <u>All Delinquents</u>

And as we have shown, this design is flawed by three uncontrolled threats to internal validity: <u>mortality</u>, <u>maturation</u>, and <u>regression</u>. The secondary finding of no difference between UDIS and DOC, on the other hand, is based on a slightly more complicated design,

0 X_1 0 <u>DOC</u>
0 X_2 0 <u>UDIS</u>
0 X_3 0 <u>Pre-UDIS</u>

which also incorporates cross-sectional comparisons. Since the UDIS and DOC groups are not randomly assigned, this design too is flawed by three (mortality, maturation, and regression) artifacts. Moreover, as this design is more complicated, we cannot easily control these threats as we did in the case of the AIR major finding.

The AIR authors found no statistically significant differences between the two groups (UDIS and DOC) by such criteria as age, family background, or seriousness of offense; but the two groups were significantly different in terms of drug use. Overall, the AIR authors concluded that, if these two groups were different at all, the UDIS delinquents were "better." Thus, any finding of a difference in post-release delinquency cannot unambiguously be attributed to the UDIS or DOC "treatments."

The AIR report found no differences in post-release delinquency rates between the two groups. Given the slight edge of the UDIS group over the DOC group, this finding could be interpreted to mean that DOC was somewhat more effective than UDIS, though the AIR report does not make this claim. The AIR analysis consisted of a comparison of UDIS and DOC at three, six, nine, and twelve months after release, with the following outcome in terms of percent rearrested or with any police contact by a given month.

	3 Months	6 Months	9 Months	12 Months
UDIS	30.4%	51.4%	61.9%	64.9%
DOC	23.0%	51.1%	62.9%	69.2%

At the end of 12 months, 35.1% and 30.8% of the UDIS and DOC delinquents respectively had not been rearrested. The AIR authors conducted a simple chi-square test which was interpreted to mean that these percentages were not statistically different from each other.

But the chi-square test is a non-parametric test and relatively weak. The difference between UDIS and DOC would have to be extreme before a statistically significant chi-square statistic could be generated.

More important, we note an unusual phenomenon at work in these figures. During the first three months, UDIS delinquents fail at a much faster rate than DOC delinquents. In successive months, however, the DOC delinquents catch up and pass the UDIS delinquents. One must speculate about other lengths of follow-up, say 15 months. Would the DOC delinquents continue to fail at a faster rate? If this were the case, the difference between DOC and UDIS would grow larger with time.

In Table 3.0 on page 57, we first show police contact by length of follow-up for the UDIS, DOC, and PRE-UDIS Baseline groups. The most troublesome aspect of these data is that the length of follow-up varies

Table 3.0

Successes and Failures (Any Police Contact) by Length of Follow-up for UDIS, DOC, and Pre-UDIS Baseline Delinquents.

	UDIS		DOC		Pre-UDIS	
	F	S	F	S	F	S
Followed-up Less						
than ... 30 days	10	10	0	5	3	0
... 60 days	10	10	4	6	9	0
... 90 days	8	7	10	4	14	1
...120 days	12	6	4	4	8	4
...150 days	7	8	8	4	8	0
...180 days	4	2	3	4	6	1
...210 days	2	6	0	5	9	3
...240 days	2	6	1	7	5	1
...270 days	1	5	0	1	3	0
...300 days	0	1	1	5	1	1
...330 days	0	3	1	5	0	2
...360 days	0	5	0	5	0	1
...390 days	0	3	0	3	0	0
...420 days	0	3	1	2	2	1
...450 days	0	1	0	5	0	0
...480 days	0	2	0	1	2	2
...510 days	0	4	0	1	1	0
...540 days	0	1	0	1	1	0
...570 days	0	1	0	1	0	3
...600 days			0	2	1	0
...630 days					0	1
...660 days			0	1	0	1
...690 days					2	0
...720 days or longer					2	23

from a single day for some delinquents to over two years for others. On the face of it, this aspect of the data makes a group contrast impossible. We see, for example, that two Pre-UDIS Baseline delinquents have had a police contact two years after "treatment." The simple recidivism rates for the UDIS, DOC, and pre-UDIS groups are 40%, 32%, and 63% respectively, but it is likely that these figures would change drastically if the UDIS and DOC groups were followed-up for longer periods of time.

An alternative which is used extensively in the AIR report is to consider only those delinquents who have been followed-up for at least one year. If this convention is used, the recidivism rates are 79%, 67% and 71%. This approach has two major shortcomings. First, the statistics ignore valuable data by excluding the histories of delinquents who were not followed-up for at least one year. Second, and more important, it assumes that delinquents in all three groups are failing at the same rate with respect to time. This is clearly not the case. The data shown in Table 3.0 indicate that the single most obvious impact of the UDIS "Treatment" is a rapid rate of failure. Those UDIS delinquents who fail do so almost immediately. Ten UDIS delinquents failed on the first day after "treatment" and these ten are 18% of all the UDIS failures. In contrast, no DOC delinquent failed within 30 days of release.

In "The Mathematics of Behavioral Change" (Maltz & McCleary, 1977), a method of computing recidivism rates that controls for variance in follow-up as well as variance in the rate of failure with respect to time is described. This method assumes that there are two types of delinquents, successes (never rearrested) and failures. By analyzing the time to failure of each delinquent, the method allows us to predict the percent in each cohort expected to fail after an extremely long period of time. The results of our analysis are as follows:

(1) The UDIS "treatment" is associated with a high rate of failure, though not with a high level of failure. Fewer UDIS delinquents will have a future police contact, but, those UDIS delinquents who will have a future contact, will have that contact almost immediately.

In Figure 3.0(a), page 61, we show the expected rates and levels of failure for UDIS and Total DOC (DOC plus Pre-UDIS Baseline) samples. Cumulative failures for UDIS asymptote at 56% while cumulative failures for Total DOC asymptote at 68%. This difference is statistically significant at the 95% confidence level. The rates of failure for the two groups are .0071 and .0036 and this difference too is statistically significant.

While this contrast is meant to assess the difference in treatments (UDIS versus DOC), the contrast may not be valid. It assumes that the treatments for DOC and Pre-UDIS Baseline groups are the same, and as we will demonstrate, there is considerable evidence to refute this hypothesis.

A more valid contrast might be between UDIS and DOC as shown in Figure 3.0(b) on page 62. For UDIS, 56% of the delinquents are expected to fail at the rate of .0071. For DOC, 50% of the delinquents are expected to fail at the rate of .0042. The difference in expected failures is not statistically different, although the difference in rates is. If the ten UDIS delinquents who failed on the first day are excluded from the analysis, 50% of the UDIS delinquents are expected to fail ultimately, the same percentage as the expected failures for DOC.

We have also considered the possibility that the post-release opportunities for police contact between the two groups are not equal. DOC delinquents are subject to parole supervision after "treatment" while UDIS delinquents are not, and thus the two groups might not be expected to have the same rate of failure in the first few months post-release. DOC delinquents might be expected ceteris paribus to have a lower rate in the first few months after "treatment." This is the equivalent of hypothesizing that the DOC "treatment" does not end at the time of release and the data support this hypothesis.

Recidivism data typically show the greatest rates of failure immediately after release. The longer a delinquent is free and without a police contact, the greater the delinquent's chances of never having a police contact. DOC failures do not follow this pattern. There are more failures in the second month than in the first, and more failures in the third month than in the second. But if the data are collapsed across time, the distribution of failures looks more normal and the rates of failure for UDIS and DOC are not statistically different: .0071 versus .0068. Moreover, the expected failures in the DOC group are now 54% versus 50% for the UDIS group, also not a statistically significant difference.

By this contrast, we tentatively conclude that there is little real difference between outcomes of these two groups. It appears that the UDIS delinquents can be expected to have fewer police contacts in the long run than the DOC group, although those UDIS delinquents who will have contacts will do so rather quickly after release. It is important to remember in interpreting these data that the youths were not randomly assigned to UDIS and DOC.

(2) Both UDIS and DOC have lower recidivism rates than the Pre-UDIS Baseline group. This difference is both statistically and substantively significant, and in one sense, is the biggest surprise of our analysis.

In Figure 3.0(c) on page 63, we show the expected failures over time for the combined UDIS and DOC samples versus the Pre-UDIS Baseline sample.

FIGURE 3.0(A)

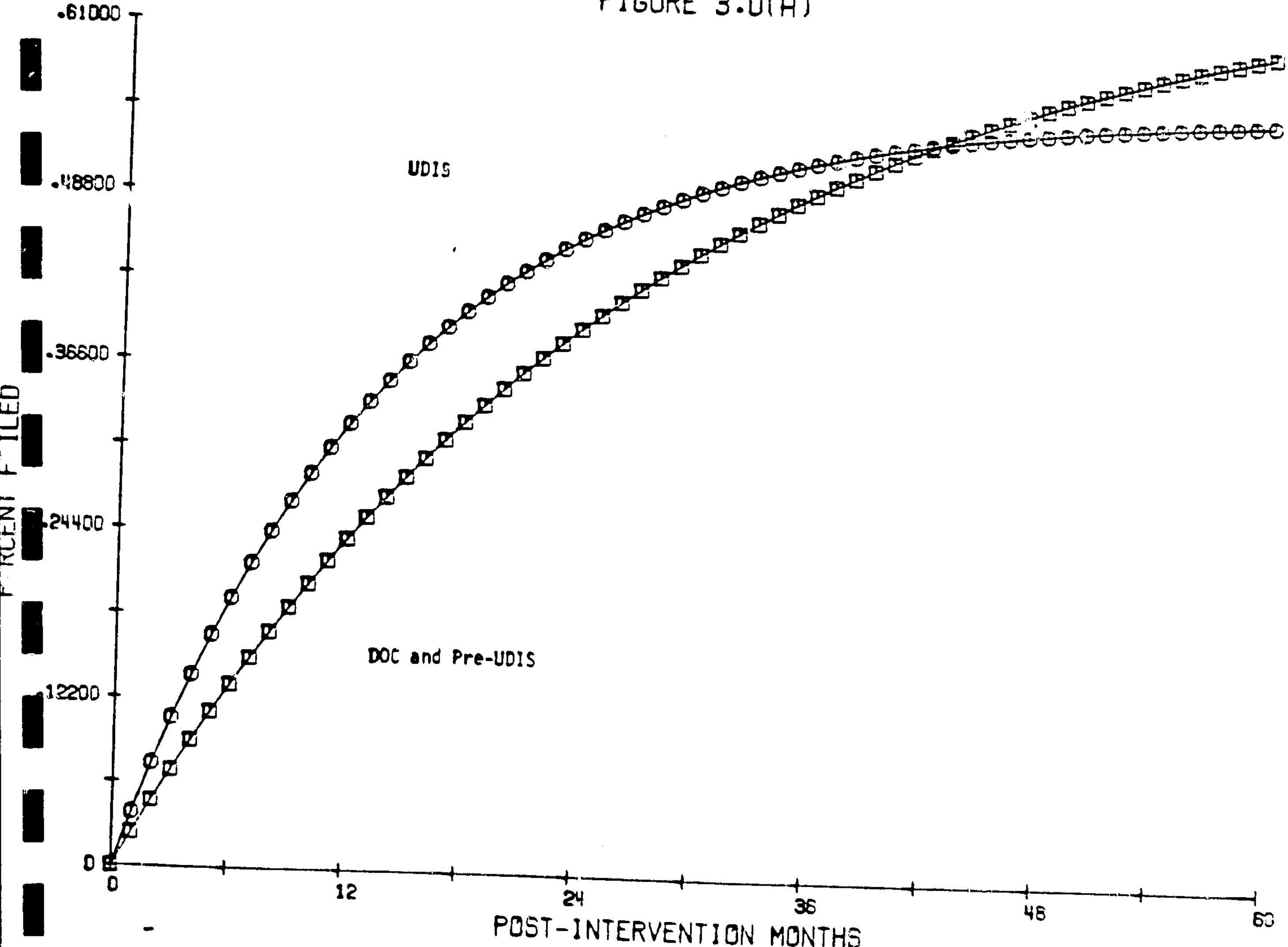

.61000
.48800
.36600
.24400
.12200
0
PERCENT F ILED
UDIS
DOC and Pre-UDIS
0
12
24
36
48
60
POST-INTERVENTION MONTHS

FIGURE 3.0(B)

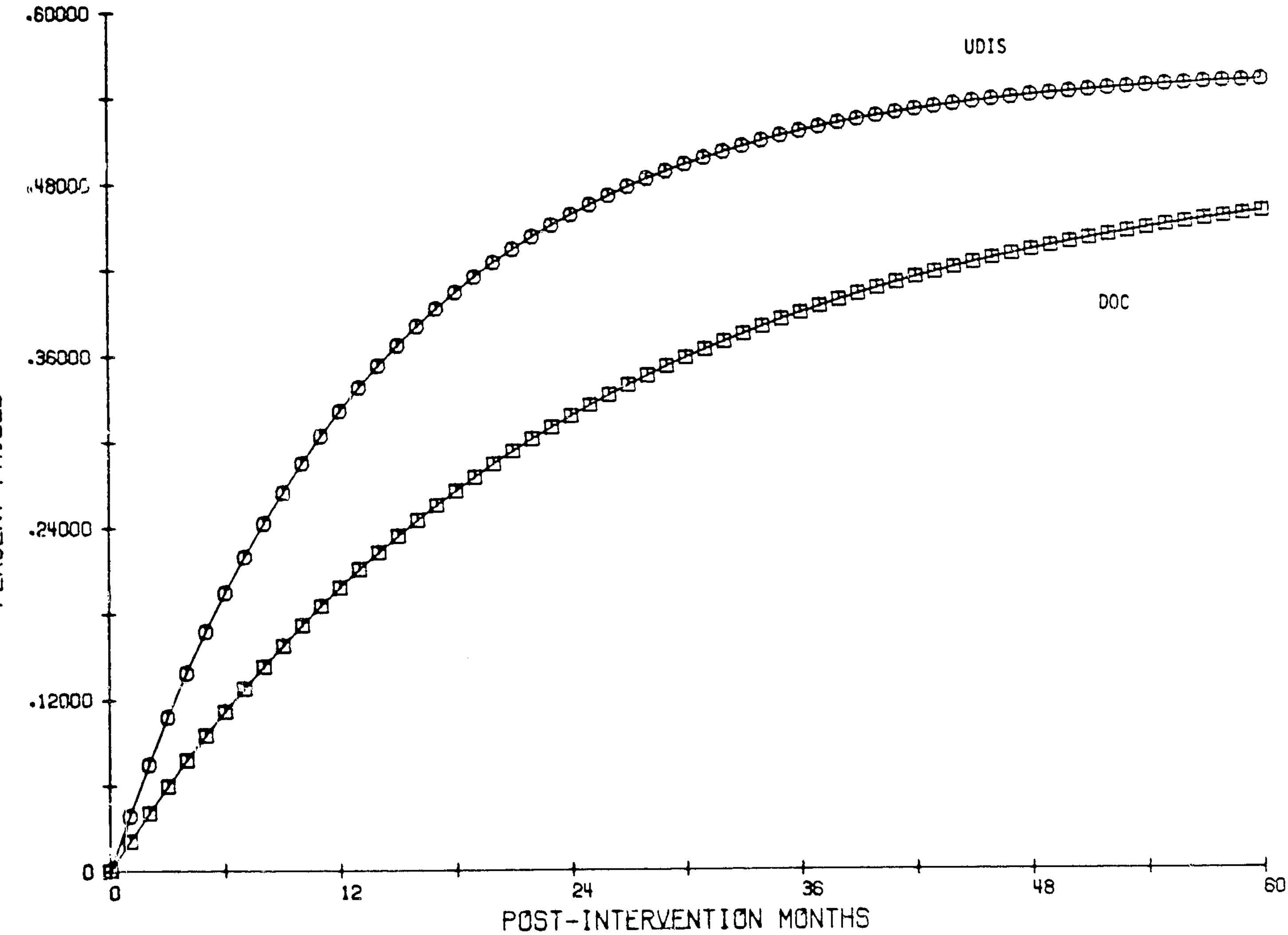
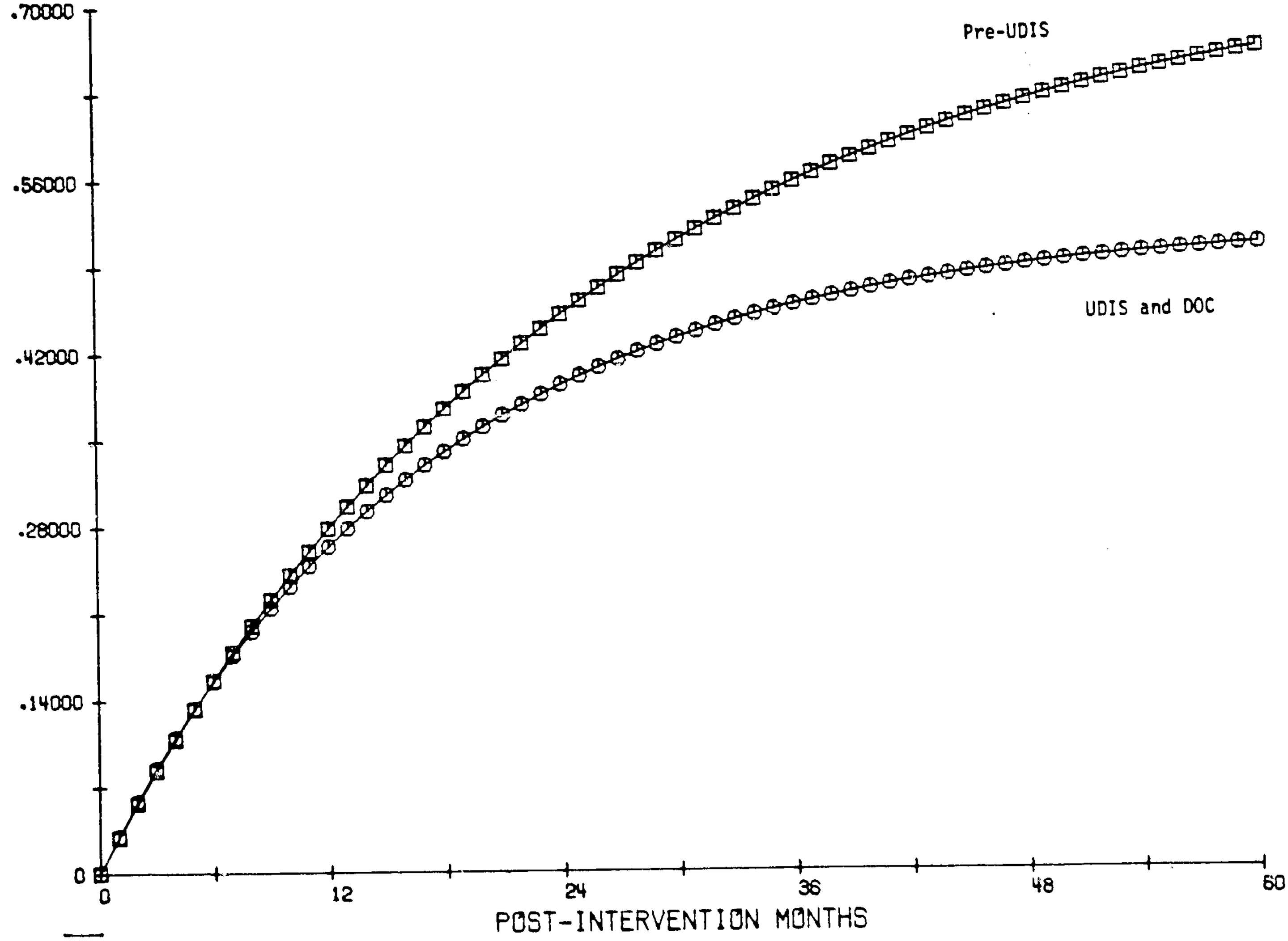

.60000
.48000
.36000
.24000
.12000
0
PERCENT FAILED
UDIS
DOC
0
12
24
36
48
60
POST-INTERVENTION MONTHS

FIGURE 3.0(C)


.70000
.56000
.42000
.28000
.14000
0
Pre-UDIS
UDIS and DOC
0
12
24
36
48
60
POST-INTERVENTION MONTHS

For the combined UDIS and DOC samples, 52% of the delinquents are expected to fail at a monthly rate of .0059. For the Pre-UDIS Baseline sample, 73% of the delinquents are expected to fail at a monthly rate of .004. In terms of the percentages ultimately failing, the UDIS and DOC groups have over 40% fewer failures than the Pre-UDIS Baseline group.

To what can this highly significant difference be attributed? Considering only these data, the climate for corrections has apparently changed for the better with regard to recidivism at the time of the creation of UDIS, and one is tempted to attribute that efficacious shift to the presence of UDIS. But we must remember that our data are by definition separated into "During UDIS" (UDIS and DOC) and "Before UDIS" (pre-UDIS baseline) categories. Other shifts in corrections policy or practice which overlap with these distinctions may be equally "causal." In our preliminary exploration of these matters, it appears that court intake and probation policy, for example, changed at a time which is confounded with our During UDIS and Before UDIS categories. In addition, there has been a secular trend downward in the numbers of youths incarcerated in DOC. The trend is acute in Illinois but apparent in other states as well. This trend may have resulted in more adequate treatment for those who remain in prison and for those in alternatives as well. This, too, is confounded with our time-related categories. At this point, we can only conclude that there has been a chronological shift corresponding in some degree to the creation of UDIS, and that the shift is toward fewer post-incarceration failures.

4.0 Within-Program Contrasts: "UDIS Levels" and the Appropriateness of AIR Evaluation of UDIS

Most of this paper has been devoted to the "suppression effect" which dominates both the AIR evaluation and all summaries of that evaluation. We have demonstrated that this major finding can be explained as the inevitable result of statistical artifacts.

Another portion of their evaluation concerns the relative effectiveness of various UDIS programs in "suppressing" delinquency. This portion of the AIR evaluation was based on the design

UDIS Level I	0	X_1	0
UDIS Level II	0	X_2	0
UDIS Level III	0	X_3	0

where the UDIS "Levels" are groups of UDIS programs supposedly varying in "severity." UDIS programs included in Level I are the least "severe" (group homes, for example) and UDIS programs included in Level III are the most "severe" incarceration, for example). On the basis of contrasts among the three Levels of UDIS, the AIR authors concluded that the more intensive the UDIS placement, the more profound the suppression effect. This finding seems to support their major "finding" of a suppression effect.

Our primary concern here will be with the arbitrary classification of UDIS programs into Levels I, II, and III. Before we address this issue, however, we should note that the claimed relationship between UDIS Level and suppression is spurious.

First, the regression results which we reported earlier indicate that Time Served in institution or program was not related to Suppression. If there were any evidence for a general deterrent effect at all, we would expect it to show up here. But delinquents who were kept in institutions for only a short period of time tend to experience greater suppression effects than delinquents who were kept in institutions for longer periods of time. This difference is not statistically significant and should not be over-interpreted, but it nevertheless argues against any general deterrent effect such as the one claimed by the AIR report for UDIS Levels.

Second, the AIR claim of differential effectiveness for Levels I, II, and III is based on the same quasi-experimental design used elsewhere. As such, all three statistical artifacts (mortality, maturation, and regression) apply. Our earlier cautions about plausible alternative explanations pertain to this part of the AIR evaluation also.

Third, and finally, only 34 UDIS delinquents were free for one year or more in the post-intervention period. The AIR finding of differential effectiveness for Levels I, II, and III is thus based on only a handful of delinquents.

We wish to emphasize another issue as well: The manner in which this finding of differential effectiveness was achieved evidences a general lack of fit between the AIR evaluation strategy and UDIS, the program they were to evaluate. We believe this lack of fit burdens their whole report, and to clarify our concern we need briefly to describe salient aspects of UDIS.

UDIS was created to provide an innovative addition to the mix of strategies for serious juvenile offenders. UDIS was to utilize case managers who would have primary responsibility for youths referred to UDIS. Following detailed individual assessments, UDIS was to design an intervention strategy tailored to each youth. These assessments and the composite experiences of UDIS administrators and staff were to result in purchase-of-service contracts to guarantee the appropriate range and locale of services.

From this service mix, which was to be modified in response to the clientele, a placement or series of placements was to be designed to match individual client needs. The primary on-going responsibilities of each UDIS case manager were (a) to monitor the progress of the youth in the placement(s), and (b) to monitor the effectiveness of the placements in delivering the contracted services. First Placement is the term we use to designate the initial placement to which any youth was assigned. These First Placements are usually but the first in a series of UDIS placements, as described below.

In their comparisons of UDIS programs with one another as "suppressors" of recidivism, the AIR evaluation is plagued not only with the artifacts which we document in the body of this paper, but also with their use of First Placement as their only variable to distinguish the careers of UDIS youths. There are several reasons why First Placement is inappropriate as a sole sorting criterion:

(1) Sometimes a youth was in a particular First Placement, not because UDIS personnel assessed it as most appropriate, but because a judge insisted on it as a condition for referral to UDIS; (2) For some youths UDIS records indicate that everyone agreed that an alternative First Placement would be better, but that the more appropriate placement was temporarily full; (3) Sometimes a youth would fail to make anticipated progress in a First Placement and would be moved into another placement (perhaps the capacity to recognize a youth failure or placement failure should be considered a possible strength of UDIS to be thoroughly evaluated); (4) For some youths a placement was never proposed alone, but either (a) as one of two or more concurrent placements, initially conceived as a package, or (b) as but the first (and often very brief) step in a pre-arranged package of sequenced placements.

In short, UDIS saw its mission as supplying a flexible range of program packages finely tuned to the needs of each youth. But just as recidivism, and recidivism alone, became the only dependent variable carefully scrutinized by AIR in their evaluation of UDIS, so First Placement, and First Placement alone, became the sole independent variable to distinguish the careers of UDIS youths. By limiting themselves to these variables, AIR severely limited what could be known about the viability of the UDIS approach, and in addition provided very few data of administrative utility.

Even though we consider the selection of recidivism by itself (however measured) to be myopic perspective of what UDIS was trying to accomplish, and First Placement to be a highly misleading way to characterize UDIS placement careers, our quantitative analysis is necessarily limited to responding to what AIR chose, because that is all that is available on the data tape purchased from AIR. Recidivism is the only outcome variable they code, and First Placement is the only placement measure they provide.

There is another substantial problem in the First Placement data available to us: the AIR data tape does not distinguish individual placements, or even types of placements (e.g., group homes, intensive care, foster homes, etc.) Instead we are limited to AIR's coding of First Placement into three virtually arbitrary "levels." Levels I, II, and III are supposed to represent increasingly drastic interventions, but in their aggregation AIR lumps together placements of widely varying types and placements which attempted to impact youths in fundamentally different ways. Level III placements, for example, included both wilderness-stress programs and high security therapeutic hospitalization. Similarly, programs with vastly different reputations (but bearing the same descriptive types, largely for purposes of remuneration -- e.g., various stripes of "group homes") have been equated with one another.

It may be highly misleading to equate programs within such arbitrary categories. While wilderness-stress programs and therapeutic hospitalization programs can be conceived of as "intensive" on some dimension, they are quite different and should not casually be collapsed. And for analytic purposes it may be just as misleading to label a group of placements simply because they share a nominal category (e.g., "group home") as it is to label a group of delinquents simply because they share a nominal category (e.g., "chronic offender"). Sample sizes may have dictated some of AIR's collapsing decision, but this purely statistical issue should not misdirect policy.

The UDIS staff made important distinctions among the various placements in terms of "quality." For example, one Level III placement was especially well regarded by the UDIS case managers and the dedicated staff of this program might well have been highly successful wherever they worked. Since a high proportion of the youths in Level III placements were in that single highly regarded program, the "quality" of program personnel may be profoundly intertwined with the "Level III" category. The AIR authors note the high ratings by UDIS case managers of the Level III placements, but in general did not deal with "quality" of placement as it relates to their recidivism outcomes. We feel that the interchangeability of placements within a nominal category is an empirical issue and not merely to be assumed.

For these reasons we have not pursued the issue of the relative effectiveness of UDIS placements any further. Because we have shown that AIR's "suppression effect" washes away; because First Placement is a gross characterization of UDIS placement careers; and because we are even more limited by AIR's three categories of First Placement, it does not seem useful to re-analyze those data.

This is not to say that within-program contrasts were not called for by the evaluation task. Indeed, within-program contrasts (e.g., between workers; between chronological periods in the life of UDIS; between program packages) would have been the most useful contrasts of all. But the design selected by AIR and the data available from them do not allow these contrasts.

5.0 APPENDIX

DRAFT:

Do not cite or quote

without permission

ARTIFICIAL INFLATION OF A POISSON RATE BY A "SELECTION EFFECT"

by

Michael D. Maltz
University of Illinois at Chicago Circle

Stephen M. Pollock
University of Michigan
Ann Arbor

July, 1978

ACKNOWLEDGEMENTS

This research was supported by the National Institute of Law Enforcement and Criminal Justice under grant number US LEAA 77 NI 99-0073. Computing services were provided by the UICC Computer Center; their assistance is gratefully acknowledged.

ARTIFICIAL INFLATION OF A POISSON RATE
BY A "SELECTION EFFECT"

Figure 1 is taken from an evaluation of the Illinois Unified Delinquency Intervention Services (UDIS) Program [Ref. 3]. It shows empirical rates of police contacts obtained from three different cohorts of juveniles, as a function of time before (and after) entry into their three respective correctional programs. As can be seen, the number of police contacts per 100 juveniles per month increases appreciably until the time of selection (i.e., sentencing to a correctional program), after which the juveniles' behavior appears to moderate considerably. MURRAY et al. [3] interpret this in terms of a "suppression effect," and conclude that it is independent of program type because the observed pattern of the rate of police contacts is the same for the three cohorts depicted in Figure 1. The implication is that the rate of contacts is "suppressed" just due to the juveniles' having been sentenced to a program.[1]

We show here how this effect may be just as easily attributed to a "selection artifact" caused by selection rules used by juvenile judges when sentencing juveniles to correctional programs. This artifact inflates the true police contact rate prior to intervention. In particular, we show how a selection rule[2] of

> If a youth has just had a police contact, and if he has had k prior police contacts the last τ months, sentence him to a correctional program

can cause the type of relationship shown in Figure 1, even if the youths in question have a <u>constant</u> rate of police contacts over time. The steep rise for the constant-rate cohort is caused by the aggregation of the juveniles' most active-time periods in the epoch just prior to selection: it is the selection rule and not a true increase in the delinquency of the juveniles that causes the rise.

We can also show that a slightly more complicated (but still stationary rate) model of juvenile behavior will produce the same pattern of contact rate vs. time.

I. THE EFFECT OF A SELECTION RULE

Consider an individual's "events" (i.e., police contacts in the case at hand) to occur according to a stationary Poisson process with constant rate λ. By definition, for each individual the times between successive events are independent random variables, each with the same probability density function:

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & 0 \leq t < \infty \\ 0 & t < 0 \end{cases} , \tag{1}$$

and if we fix a time interval of length x, then the probability that exactly n events occur in that interval is $\frac{(\lambda x)^n}{n!} e^{-\lambda x}$.

Now consider a large number of individuals, each with the same event rate λ, and suppose that we select some subset of individuals for observation based on their past history. In particular, suppose we select individuals if and only if

a) An event has just occurred, and

b) k other events have previously occurred within a time period of length τ $(k \geq 1)$.

(In our particular case the "event" is a police contact and the "selection" is sentencing to a correctional program.)

It is now possible to determine the rate of occurrence of events for the individuals who have thus been selected, for the time period prior to the selecti conditioned upon the fact that they have been selected. As we shall see, this rate depends on the selection process and is in fact not λ.

If we set $t = 0$ to be time of occurrence of the intervention event, we need to compute the rate for time $t < 0$. However, since the Poisson process is reversible, the analysis remains the same if the time axis is reversed. The selection process is thereby equivalent to one which selects an individual if and only if

a) An event has just occurred, and

b) The next k events occur within a time period of length τ.

We now need to compute $r(t|\tau)$, the rate of occurrence of events at time t, ($t > 0$), given that an event occurs at $t = 0$ and that k more events occur in the interval $0 < t \leq \tau$. By the conventional definition of a conditional rate

$$r(t|\tau)dt \equiv Pr[\{E(t)\}|T_k \leq \tau] \tag{2}$$

where

$\{E(t)\}$ is the event {police contact between t and t+dt}

and

T_k is the time of occurrence of the kth event following the one at $t = 0$

Using Bayes' theorem, equation (2) can be rewritten as

$$r(t|\tau)dt = \frac{Pr[T_k \leq \tau|\{E(t)\}] \cdot Pr[\{E(t)\}]}{Pr[T_k \leq \tau]} \tag{3}$$

Since the events occur according to a Poisson process, we have

$$Pr[T_k \leq \tau] = 1 - \sum_{i=0}^{k-1} \frac{(\lambda\tau)^i}{i!} e^{-\lambda\tau} \tag{4}$$

and

$$\Pr[\{E(t)\}] = \lambda dt \tag{5}$$

The first term in the numerator of (3) can be evaluated separately for $0 < t \leq \tau$ and $\tau < t < \infty$. In the former case the condition $\{E(t)\}$ puts at least one event in the interval of length τ, leaving k-1 events to be accounted for. Thus for $0 < t \leq \tau$

$$\Pr[T_k \leq \tau | \{E(t)\}] = \Pr[T_{k-1} \leq \tau] \tag{6}$$

If $t > \tau$ the condition $\{E(t)\}$ has no effect on events occurring at times before τ. Therefore, for $\tau < t < \infty$,

$$\Pr[T_k \leq \tau | \{E(t)\}] = \Pr[T_k \leq \tau] \tag{7}$$

Substituting (5) through (7) into (3) results in

$$r(t|\tau) = \begin{cases} \dfrac{\lambda \Pr[T_{k-1} \leq \tau]}{\Pr[T_k \leq \tau]} & 0 < t \leq \tau \\ \lambda & \tau < t < \infty \end{cases} \tag{8}$$

For k = 1 the event rate becomes, using equations (4) and (8),

$$r(t|\tau) = \begin{cases} \lambda/[1-e^{-\lambda\tau}] & 0 < t \leq \tau \\ \lambda & \tau < t < \infty \end{cases} \tag{9}$$

When the time axis is re-reversed, this result shows that the selection procedure alone makes it appear that there is an increase in the rate during the time interval between $-\tau$ and 0 (Figure 2). Since by assumption the intervention has no effect on the individuals, the event rate reverts to λ for all time <u>after</u> intervention.

Of course, the step rises in rate shown in Figure 2 show no great resemblance to data shown in Figure 1, although it does show an increase just prior to $t = 0$. The next two sections show how this can be obtained from equation (8) when certain conditions (i.e., τ fixed, all offenders meeting the criterion sentenced) are relaxed.

Effect of a Distribution on τ

Equation (8) was derived for a fixed value of the time interval τ, the time within k+1 events must occur for selection. But not all judges are expected to use the same τ -- some may use six months, some two years. Let us assume that among all judges the interval τ has a cumulative frequency distribution $G(\tau)$; i.e., the fraction of intervals less than or equal to x is $G(x)$. The unconditional rate of occurrence of events at (reversed) time t then becomes

$$r(t) = \int_0^\infty r(t|\tau)dG(\tau) = \lambda\left[\int_0^t dG(\tau) + \int_t^\infty \frac{\Pr[T_{k-1}<\tau]}{\Pr[T_k<\tau]}dG(\tau)\right] \qquad (10)$$

This has the appropriate smooth behavior for all differentiable $G(\tau)$. For example, Figure 3 gives the rate $r(t)$ for $G(\tau)$ uniform between 0 and 24 months,[3] for different values of k, superimposed on the data.[4]

Effect of Selection Probability

Thus far we have assumed that the selection rule is invariably followed, that any juvenile generating k events in a time interval τ will be selected for intervention. In contrast, let us now suppose that a judge bases his decision only on the time since last contact. In particular, suppose a judge has a probability $p(x)$ of selecting a juvenile for a program, where x is the time since his last contact. Then we have

$$r(t)dt = \Pr[\{E(t)\}|\text{select}]$$

$$= \frac{\Pr[\text{select}|\{E(t)\}]\Pr[\{E(t)\}]}{\Pr[\text{select}]}$$

where

$$\Pr[\{E(t)\}] = \lambda dt,$$

$$\Pr[\text{select}] = \int_0^\infty p(x)\lambda e^{-\lambda x}dx$$

and

$$\Pr[\text{select}|\{E(t)\}] = \Pr[\text{select} \cap T_1 < t|\{E(t)\}]$$

$$+ \Pr[\text{select} \cap T_1 = t|\{E(t)\}]$$

$$+ \Pr[\text{select} \cap T_1 > t|\{E(t)\}]$$

$$= \int_0^t p(x)\lambda e^{-\lambda x}dx + p(t)e^{-\lambda t} + 0$$

so that

$$r(t) = \lambda \left[\frac{\int_0^t p(x)\lambda e^{-\lambda x}dx + p(t)e^{-\lambda t}}{\int_0^\infty p(x)\lambda e^{-\lambda x}dx} \right] \qquad (11)$$

[Note that if we let $p(x) = 1$ for x between 0 and τ and $p(x) = 0$ elsewhere, the solution to equation (11) is equation (9), as expected.]

Equation (11) thus provides an alternative or complementary explanation to the inflation of the fundamental Poisson rate. The Appendix shows how these two models may be combined.

II. A MARKOV MODEL

A slightly different model can also serve to explain the data of Figure 1. This model is based on the behavior of the offenders, while the previous ones were based on the behavior of the judges. It too can explain a rise in the police contact rate even though the underlying process is stationary.

We can define two behavioral states for a juvenile:

State 1 ("active"), in which he has a police contact rate of λ_1;

State 2 ("quiescent"), in which he has a police contact rate of λ_2 ($\lambda_2 < \lambda_1$).

Let transitions between these states be described by a continuous time Markov process, with α and β the transition rates from states 1 to 2, and 2 to 1, respectively. Again, we consider a large number of individuals, each undergoing transitions (with the same rates) and having police contacts (with the same state-specific rates).

In this model, the police contact event itself becomes the source of a "selection bias," so we need not consider a critical time period τ, or a selection probability $p(x)$. Instead, let us assume that the observation of a youth in the cohort begins at some time independent of the number of police contacts he has generated and independent of his present state (for example, when he reaches age 13). Let us further assume that once any police contact is generated after this time, he is selected (i.e., sentenced to a correctional program).

Defining $t = 0$ to be the time of selection, and

$\{S(t)=i\}$ = event {state of person is i at time t}; i = 1 or 2,

$\{E(t)\}$ = event {police contact between t and t+dt}

We need to compute:

$$\begin{aligned} r(-t)dt &= \text{event rate at } -t \text{ given selection at } t = 0 \\ &= \Pr[\{E(-t)\}|\{E(0)\}] \\ &= \Pr[\{E(-t)\}|\{S(-t)=1\},\{E(0)\}]\Pr[\{S(-t)=1\}|\{E(0)\}] \\ &\quad + \Pr[\{E(-t)\}|\{S(-t)=2\},\{E(0)\}]\Pr[\{S(-t)=2\}|\{E(0)\}] \end{aligned}$$

which, by definition of λ_1 and λ_2, gives

$$r(-t) = \lambda_2+(\lambda_1-\lambda_2)\Pr[\{S(-t)=1\}|\{E(0)\}] \tag{12}$$

The last term in (12) is the probability that a person was in state 1 t time units ago, given a contact is made now. This can be determined using

Bayes' rule:

$$\Pr[\{S(-t)=1\}|\{E(0)\}] = \frac{\Pr[\{E(0)\}|\{S(-t)=1\}]\Pr[\{S(-t)=1\}]}{\Pr[\{E(0)\}]} \tag{13}$$

If the selection program commenced when the "system" of individuals was in steady state, then

$$\Pr[\{S(-t)=1\}] = \text{steady-state probability that the system is in state 1}$$

$$= \frac{\beta}{\alpha+\beta} \tag{14}$$

and

$$\Pr[\{E(0)\}] = \lambda_1 \left(\frac{\beta}{\alpha+\beta}\right) + \lambda_2 \left(\frac{\alpha}{\alpha+\beta}\right) \tag{15}$$

The conditional probability in the numerator of (13) may be readily obtained from the general transient solution for a two-state process:

$$\begin{aligned}
\Pr[\{E(0)\}|\{S(-t)=1] &= \Pr[\{E(0)\}\cap\{S(0)=1\}|\{S(-t)=1\}]+\Pr[\{E(0)\}\cap\{S(0)=2\},\{S(-t)=1\}] \\
&= \Pr[\{E(0)\}|\{S(0)=1\},\{S(-t)=1\}]\Pr[\{S(0)=1\}|\{S(-t)=1\}] \\
&+ \Pr[\{E(0)\}|\{S(0)=2\},\{S(-t)=1\}]\Pr[\{S(0)=2\}|\{S(t)=1\}] \\
&= \lambda_2+(\lambda_1-\lambda_2)\Pr[\{S(0)=1\}|\{S(-t)=1\}] \\
&= \lambda_2+(\lambda_1-\lambda_2)\Pr[\{S(t)=1\}|\{S(0)=1\}] \\
&= \lambda_2+(\lambda_1-\lambda_2)\left[\frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t}\right]
\end{aligned} \tag{16}$$

where now t represents time units into the past.

Substituting (13) through (16) into (12) gives, finally,

$$r(t) = \frac{\lambda_1\beta+\lambda_2\alpha}{\alpha+\beta} + \frac{\alpha\beta(\lambda_1-\lambda_2)^2}{(\alpha+\beta)(\lambda_1\beta+\lambda_2\alpha)} e^{-(\alpha+\beta)t} \tag{17}$$

Thus the general behavior of the <u>apparent</u> contact rate shows an exponential decay into the past, to a steady-state value $\bar{\lambda} = \frac{\lambda_1\beta+\lambda_2\alpha}{\alpha+\beta}$, in spite of the fact that each individual experiences only time independent, state-specific rates λ_1 and λ_2.

Indeed, if we did not force the sampling process to define t=0 to be at a contact, then the observed rate would be $\bar{\lambda}$ at all times, past and future.

Figure 4 shows r(t) from equation (17) for λ_1=.8, λ_2=.02, α=.063 and β=.007, which fits the data quite well.

This model has an additional feature, in that it allows a calculation of the contact rate in the future. (This is in contrast to the ones discussed earlier, which require the apparent contact rate to remain at λ after selection.) In particular, since programs result in some probability of releasing the youths in states 1 and 2, then the contact rate would exponentially climb (or decay) to $\bar{\lambda}$ after release. This behavior is seen, at least qualitatively, in the data of Figure 1.

III. CONCLUSION

This paper points out one problem inherent to using data obtained from quasi-experimental evaluations. We have shown how the drawing of inferences from such data must be tempered by the possibility of an artifact introduced by the selection of subjects. In doing so, we have introduced three possible models of juvenile police contact behavior and subsequent court actions, which all assume stationary contact rates yet which all produce an apparent increase in rates prior to selection.

The models we have described are quite simple, whereas the actual selection process may be much more complicated. For example, the decision to select a juvenile for a correctional program would doubtless be based on the seriousness of the instant offense and prior offenses, as well as on the number of prior offenses within a time interval.

We do not mean to imply that the entire "suppression effect" noted by Murray et al. [8] is attributable to an artifact, but rather that some part of it might well be artifactual. It is possible (with enough effort) to determine the selection rule (perhaps by asking judges); then to model the rule to estimate its effect on the data; and then to remove this effect; thus leaving us with the pure "suppression effect." However, not much would be gained by this strategy. Offender behavior may not actually be stationary, let alone Poisson; our choice of these constructs is to show what might be, not what is, the case. For example, if an alternative model posits an age-dependent increase in police contact rate, there would be more "suppression effect" and less artifact. Our point is that a program is often impossible to evaluate when the outcome measure used (in this case, police contact rate) is also the variable used in selecting people for the program; policy based on this type of evaluation can be dangerously misleading.

APPENDIX

Suppose a judge has a probability $p(T_1)$ of selecting a juvenile for a program, where T_1 is the time since the last police contact, provided that (1) there is a contact now (at t=0) and (2) there have been k contacts within a time interval τ ($k \geq 2$). We would then have (see equation (11))

$$r(t) = \lambda \left[\frac{\int_0^t p(x)f(x)dx + p(t)[1 - F(x)]}{\int_0^\infty p(x)f(x)dx} \right] \qquad \text{(A-1)}$$

where f(x) is the probability density function (and F(x) is the cumulative distribution function) of the time since last contact, conditioned on there just having been a contact ond on there having been k prior contacts within a time interval τ. To calculate f(x) note that

$$\begin{aligned} F(x) &= P[T_1 \leq x | \{\text{selection}\}] \qquad , 0 \leq x < \tau \\ &= P[T_1 \leq x | T_k \leq \tau] \\ &= 1 - P[T_1 > x | T_k \leq \tau] \\ &= 1 - \frac{P[T_k \leq \tau | T_1 > x] P[T_1 > x]}{P[T_k \leq \tau]} \end{aligned}$$

$$F(x) = \begin{cases} 1 - \dfrac{e^{-\lambda x} P[T_k \leq \tau - x]}{P[T_k \leq \tau]} & 0 \leq x < \tau \\ 0 & \text{elsewhere} \end{cases} \qquad \text{(A-2)}$$

and, using (4) we find that

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \left[\dfrac{P[T_{k-1} \leq \tau - x]}{P[T_k \leq \tau]} \right] & 0 \leq x < \tau \\ 0 & \text{elsewhere} \end{cases} \qquad \text{(A-3)}$$

These can be substituted into (A-1) and solved for any p(x).

NOTES

1. This "suppression effect" is being cited as justification for a "get-tough" policy toward juvenile delinquents. See, for example, Refs. [2] and [4].

2. Terry [5] shows that this selection rule has an empirical basis. His study of a midwestern court indicates that the invocation of formal procedures against a juvenile is strongly related to the number of his previous referrals to the juvenile authorities.

3. A computational note: from (4) we see that $e^{\lambda\tau}Pr[T_{k-1}\leq\tau]$ is the derivative (with respect to τ) of $e^{\lambda\tau}Pr[T_k\leq\tau]$, making the second integral in (10) simple to compute for τ uniformly distributed.

4. In a reanalysis of the original data [1] it was found that the lower contact rate from 24 to 48 months was in part due to the lack of data on the juveniles that far back, and in part due to the fact that many of the juveniles were <u>very</u> young then (ten to twelve years old). Going back only 24 months illustrates the rise in police contact rate without having to consider these (and other) biasing effects.

REFERENCES

1. McCleary, R., Maltz, M. D., McDowell, D., and Gordon, A. "How a Regression Artifact Can Make Any Delinquency Intervention Program Look Effective." In preparation.

2. Murray, C. A. Statement before the US Senate Subcommittee on Juvenile Delinquency, April 12, 1978.

3. Murray, C. A., Thomson, D., and Israel, C. B. "UDIS: Deinstitutionalizing the Chronic Juvenile Offender." American Institutes for Research, Washington, DC, January, 1978.

4. New York Times: editorial, June 19, 1978; letter to the editor, June 30, 1978.

5. Terry, R.M. "The Screening of Juvenile Offenders." Journal of Criminal Law, Criminology and Police Science 58:2, 173-181 (1967).

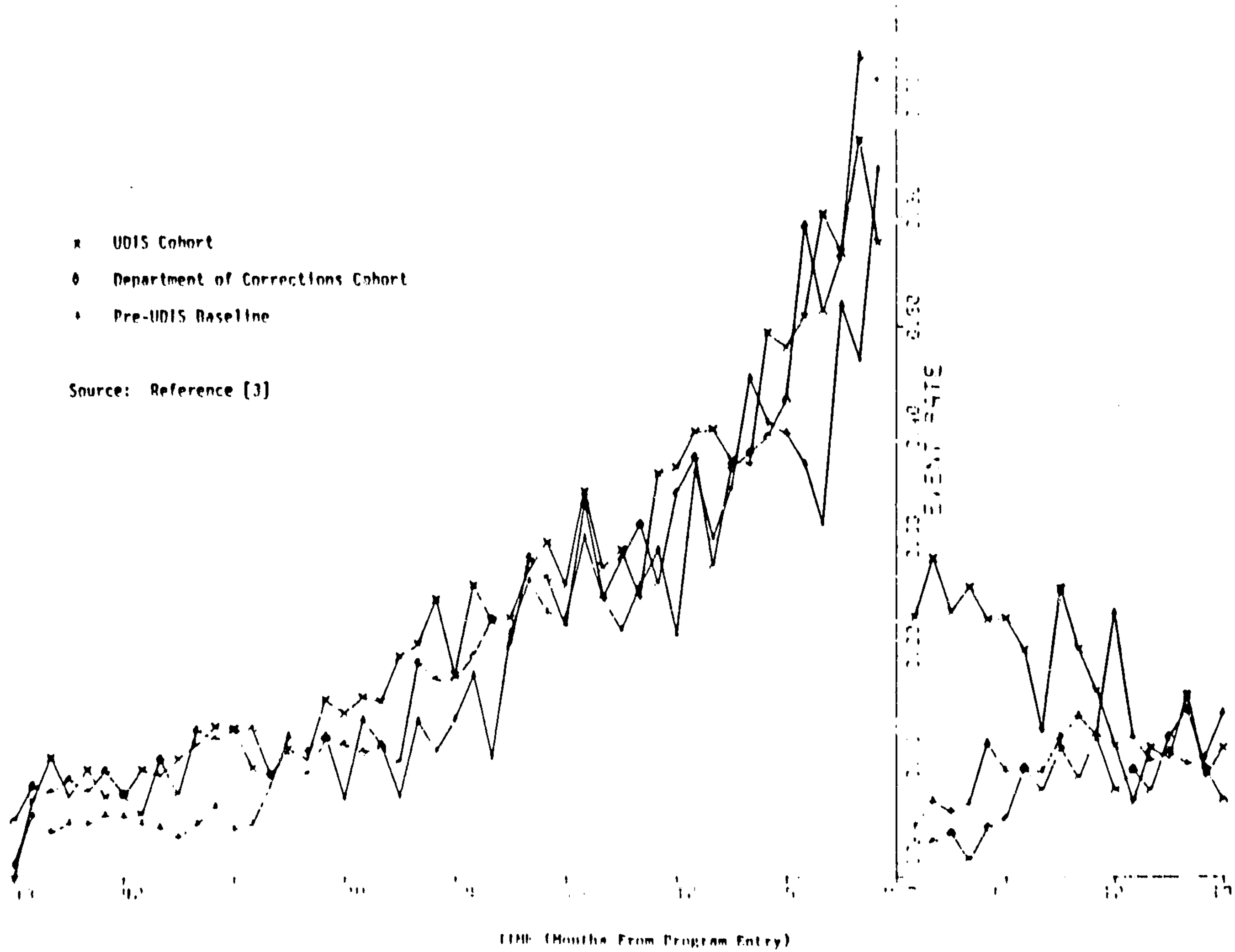


Figure 1. Monthly Rate of Police Contacts for Three Groups of Juveniles

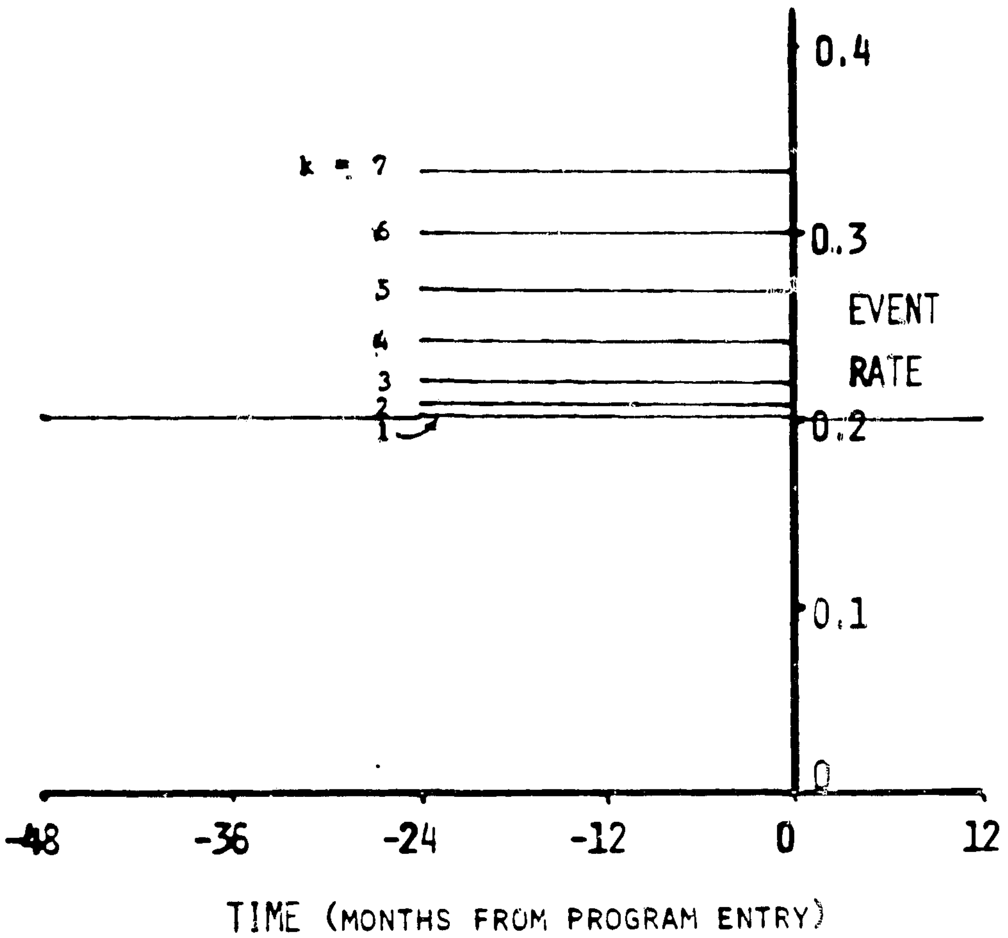


Figure 2. Event Rate Using a Selection Rule of k Events Within 24 Months Prior to Present Contact, for λ=0.2

CONTINUED

1 OF 2

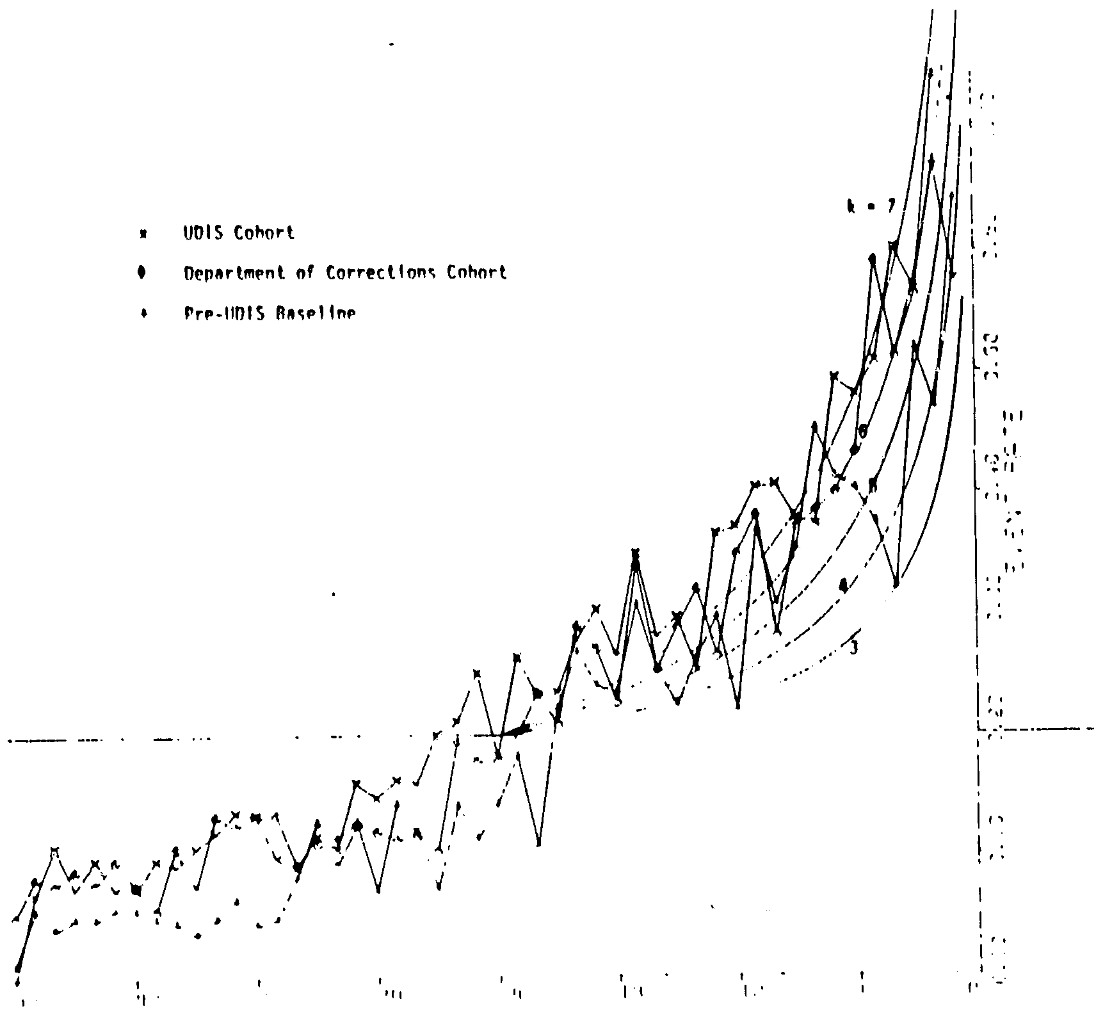


Figure 3. Event Rate When τ Is Uniformly Distributed
Between 0 and 24 Months, for λ=0.2

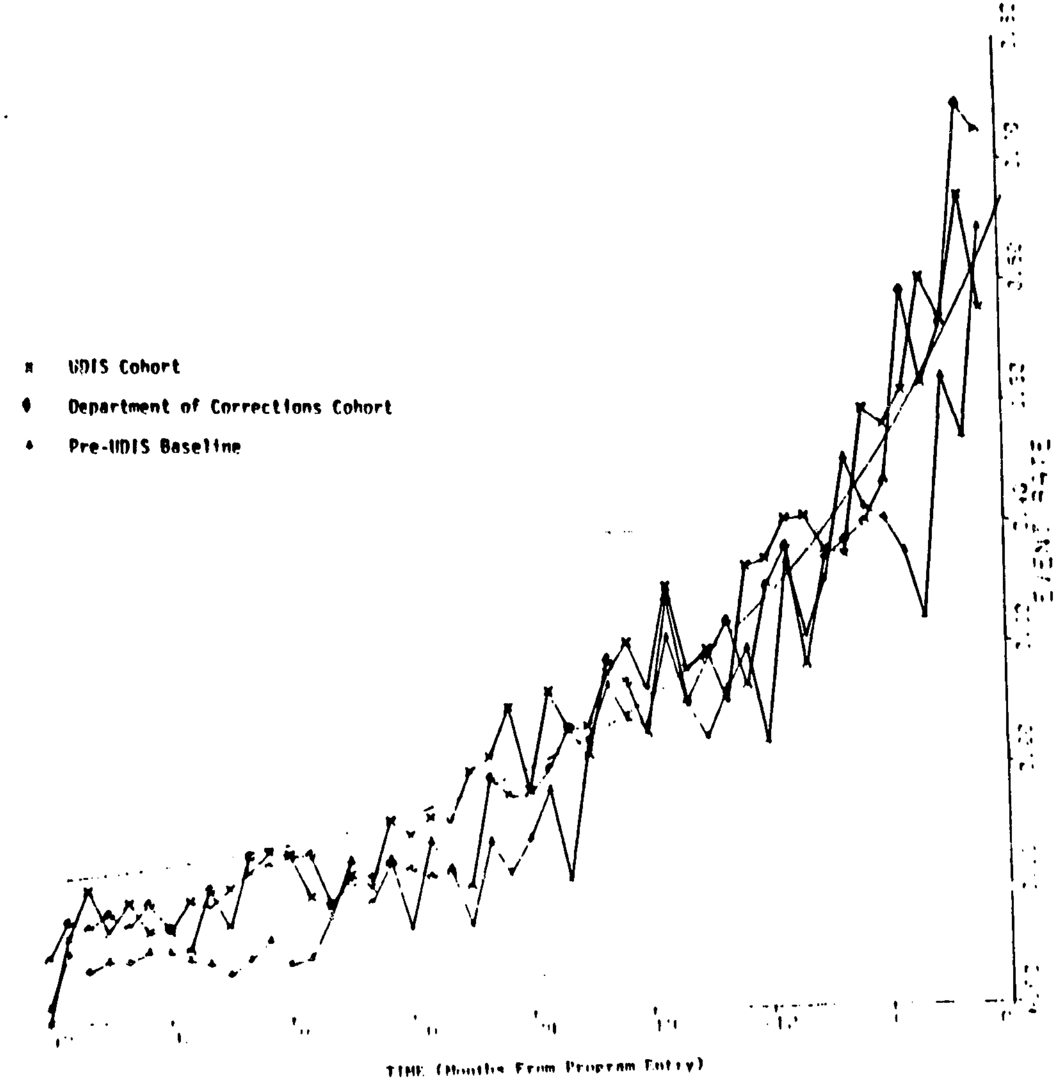


Figure 4. Event Rate Using a Markov Model

(λ_1=0.8, λ_2=0.02, α=0.063, β=0.007)