Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Artificial Inflation of a Delinquency Rate by a Selection Artifact

Michael D. Maltz, Stephen M. Pollock,

Artificial Inflation of a Delinquency Rate by a Selection Artifact

MICHAEL D. MALTZ
University of Illinois at Chicago Circle, Chicago, Illinois

STEPHEN M. POLLOCK
University of Michigan, Ann Arbor

(Received November 1978; accepted October 1979)

Cohorts of youths sentenced to a variety of correctional programs show substantial reductions in delinquent activity after leaving the programs compared to before sentencing. This paper develops models of delinquent activity and subsequent sentencing to a correctional program We show how a population of youths, whose delinquent activity is represented by a stationary stochastic process, can be selected (using reasonable selection rules) to form a cohort which has an inflated rate of delinquent activity prior to selection. When the activity rate returns to its uninflated rate after the youths are released from the program, an apparent reduction results. Based on this analysis we conclude that the reductions noted in delinquent activity may be largely due to the way delinquents are *selected* for correction rather than to the effect of the programs.

OVER THE PAST few years a number of evaluations of correctional programs have shown substantial reductions in delinquent activity when comparing delinquency rates before and after the correctional intervention. It has been claimed that the reductions are attributable to the correctional treatment administered the delinquents. We furnish an alternative explanation—that the method of selection of the cohort can artificially inflate the rate of delinquent activity just prior to selection. The subsequent return to the normal rate produces the reduction. We develop a number of models that produce this behavior.

In Section 2 we show how a constant rate of delinquent activity can be inflated if judges select delinquents for treatment based on their prior records. In Section 3 we show how similar increases can be generated by a cohort whose delinquent activity is generated by a stationary (but nonconstant rate) process. Based on these models we conclude that reductions in delinquent activity attributed to correctional treatments may be largely artifactual.

Operations Research
Vol. 28, No. 3, Part I, May–June 1980

0030-364X/80/2803-0547 $01.25

1. BACKGROUND

Figure 1 is taken from an evaluation of the Illinois Unified Delinquency Intervention Services (UDIS) Program [5]. 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

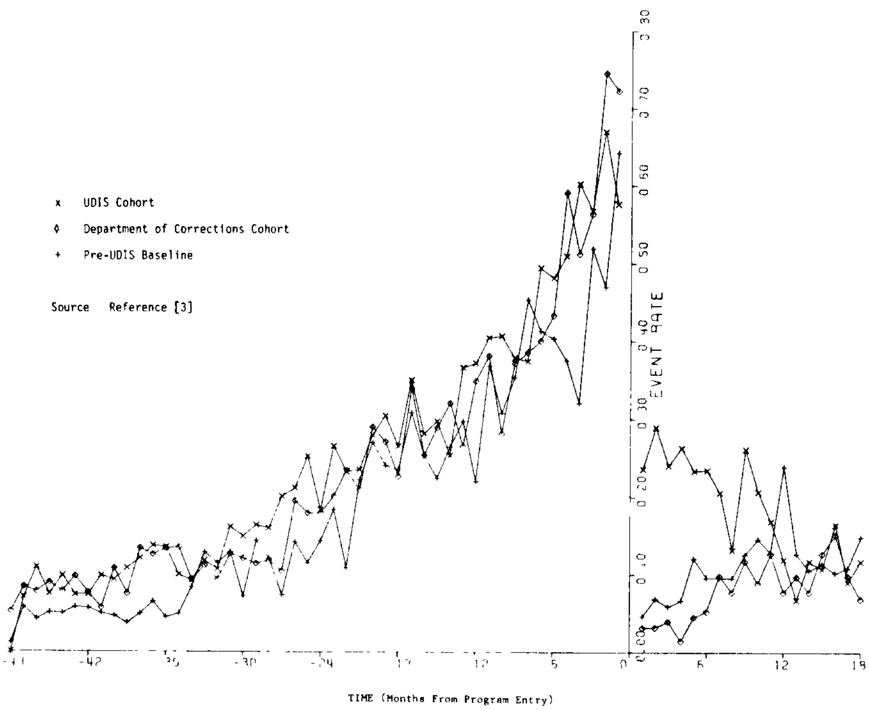

Figure 1. Monthly rate of police contacts for three groups of juveniles.

(i.e., sentencing to a correctional program), after which the juveniles' behavior appears to moderate considerably. Murray et al. [5] 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.

Other studies of delinquency programs have shown similar results. Empey and Erickson [1] and Empey and Lubeck [2] analyzed data from controlled experiments to evaluate delinquency treatments. In both cases [1, pp. 209–211] and [2, pp. 259–261] the number of arrests in the 12 months before intervention were compared to the 12 months after intervention. Reductions of about 80% were found for experimental and control. Here, too, it was inferred that experimental and control programs alike contributed significantly to reductions in delinquency, giving further support to the "suppression effect" of delinquency programs.

Such a finding has major policy implications. In particular, the suppression effect is being cited as justification for a "get-tough" policy toward juvenile delinquents; for example, see [4] and [6].

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 cohort with a *constant* rate of generation of police contacts can cause the type of relationship shown in Figure 1, if judges use this selection rule: *If a youth has just had a police contact, and if he has had at least k prior police contacts in the last* τ *months, sentence him to a correctional program.*

Terry [7] 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.

We do not mean to imply that the youths in the actual cohorts do generate a constant rate of delinquent activity. Rather, we point out here that the reason for the steep rise in Figure 1 is not necessarily because they are more active. The steep rise for the constant-rate cohort may be caused by the aggregation of the juveniles' most active time periods in the epoch just prior to selection: the rise may be caused by the *selection rule* and not by a true increase in delinquent activity by the juveniles.

We also show (in Section 3) that a slightly more complicated (but still stationary rate) model of delinquent behavior will produce the same pattern of contact rate vs. time.

2. 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 $((\lambda x)^n/n!)e^{-\lambda x}$.

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) at least k other events ($k \geq 1$) have previously occurred within a time period of length τ. (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 selection, 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. (In fact, the rate is at least k/τ, which for sufficiently large k is greater than 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 ($k \geq 1$) 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 $0 > t > \tau$. By definition, this rate is

$$r(t, \tau) \equiv \lim_{\Delta t \to 0} \Pr[\{E(t)\} \mid T_k \leq \tau]/\Delta t \qquad (2)$$

where $\{E(t)\}$ is the event {police contact between t and $t + \Delta t$} and T_k is the time of occurrence of the kth event following the one at $t = 0$. Using Bayes' theorem, the term on the right hand side of equation (2) can be rewritten as

$$\Pr[T_k \leq \tau \mid \{E(t)\}] \cdot \Pr[\{E(t)\}]/(\Pr[T_k \leq \tau]\, \Delta t). \qquad (3)$$

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

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

and

$$\Pr[\{E(t)\}]/\Delta t = \lambda + 0(\Delta t). \qquad (5)$$

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 $0 < t \leq \tau$

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

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

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

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

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

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

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

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

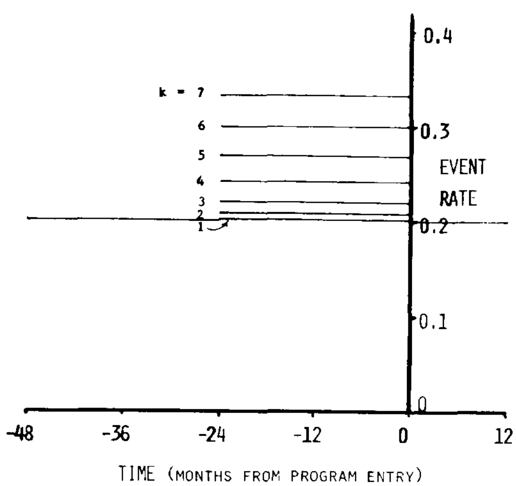

Figure 2. Event rate using a selection rule of k events within 24 months prior to present contact, for $\lambda = 0.2$.

the intervention has no effect on the individuals, the event rate reverts to λ for all time *after* 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 the value of τ used for each case may vary—due to the seriousness of the present offense, due to the offender's age, or due to judges' personal proclivities. 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

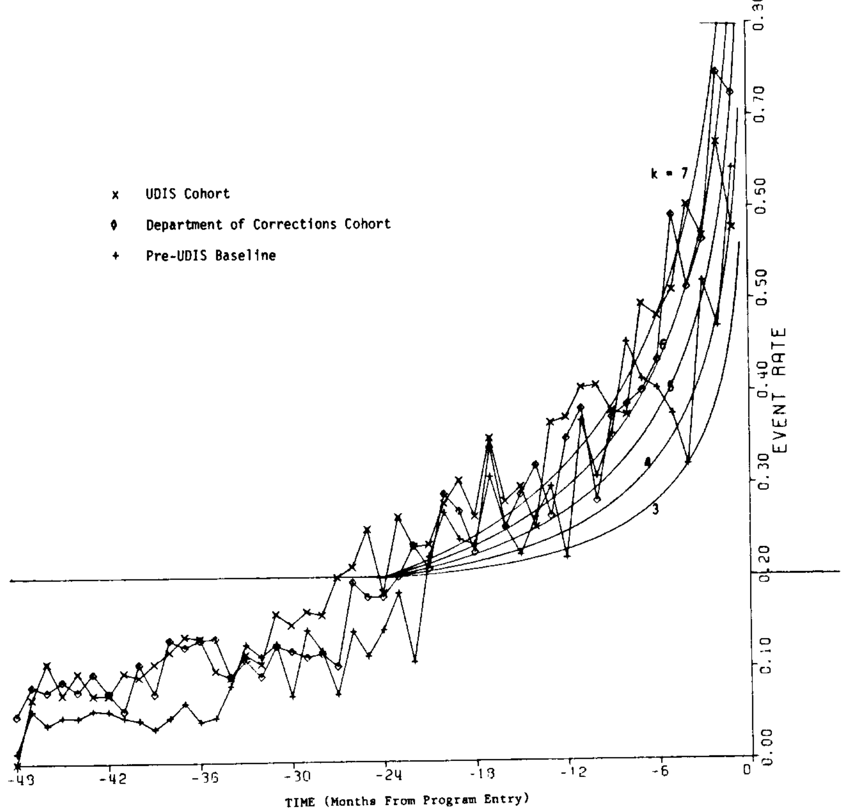

Figure 3. Event rate when τ is uniformly distributed between 0 and 24 months, for $\lambda = 0.2$.

t then becomes

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

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, for different values of k, superimposed on the data.

Certain aspects of this analysis are of interest. First, we have assumed that all cohort members have the same event rate λ, but this is not necessary: if the cohort had some distribution of event rates the result would be the inflation of all of them.

Second, it is similarly not necessary to assume that the event-generating processes are stationary. For nonstationary processes the selection rule would just change the *pattern* of nonstationarity, again upward.

Third, we assumed that τ is uniformly distributed, but this is for the sake of computational simplicity. From (4) we see that $e^{\lambda\tau}\Pr[T_{k-1} \le \tau]$ is the derivative (with respect to τ) of $e^{\lambda\tau}\Pr[T_k \le \tau]$, making the second integral in (10) simple to compute for τ uniformly distributed.

Fourth, note that our model did not fit the data from 24 to 48 months. In a reanalysis of the original data [3] it was found that the lower police contact rate during that time was in part due to the lack of data on some of the juveniles that far back, and in part due to the fact that many of them were *very* young then (10 to 12 years old). Going back only 24 months illustrates the rise in police contact rate without having to consider these (and other) biasing effects.

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

$$\begin{aligned} r(t) &= \lim_{\Delta t\to 0} \Pr[\{E(t)\}\mid \text{select}]/\Delta t \\ &= \lim_{\Delta t\to 0} \Pr[\text{select}\mid \{E(t)\}]\Pr[\{E(t)\}]/(\Pr[\text{select}]\Delta t) \end{aligned}$$

where

$$\Pr[\{E(t)\}] = \lambda\Delta t + 0(\Delta t), \quad \Pr[\text{select}] = \int_0^\infty p(x)\lambda e^{-\lambda x}\,dx$$

and

$$\begin{aligned} \Pr[\text{select}\mid \{E(t)\}] &= \Pr[\text{select} \cap T_1 < t \mid \{E(t)\}] \\ &\quad + \Pr[\text{select} \cap T_1 = t \mid \{E(t)\}] \\ &\quad + \Pr[\text{select} \cap T_1 > t \mid \{E(t)\}] \\ &= \int_0^t p(x)\lambda e^{-\lambda x}\,dx + p(t)e^{-\lambda t} + 0 \end{aligned}$$

so that

$$r(t) = \lambda \left[\int_0^t p(x)\lambda e^{-\lambda x}\, dx + p(t)e^{-\lambda t} \right] \Big/ \int_0^\infty p(x)\lambda e^{-\lambda x}\, dx. \quad (11)$$

(Note that, if we let $p(x) = 1$ for x between 0 and τ and $p(x) = 0$ elsewhere, equation (11) reduces to 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.

3. 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, $\{S(t) = i\}$ = event {state of person is i at time t}; i = 1 or 2, and $\{E(t)\}$ = event {police contact between t and $t + dt$}. We need to compute:

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

which, by definition of λ_1 and λ_2, gives

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

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:

$$\Pr[\{S(-t) = 1\} \mid \{E(0)\}] = \Pr[\{E(0)\} \mid \{S(-t) = 1\}]\Pr[\{S(-t) = 1\}]/\Pr[\{E(0)\}]. \quad (13)$$

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

$$\Pr[\{S(-t) = 1\}] = \text{steady-state probability that the system is in State 1} = \beta/(\alpha + \beta) \quad (14)$$

and

$$\Pr[\{E(0)\}] = [\lambda_1(\beta/(\alpha + \beta)) + \lambda_2(\alpha/(\alpha + \beta))]\, dt. \quad (15)$$

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

$$\begin{aligned}
\Pr[\{E(0)\} \mid \{S(-t) = 1]
&= \Pr[\{E(0)\} \cap \{S(0) = 1\} \mid \{S(-t) = 1\}] \\
&\quad + \Pr[\{E(0)\} \cap \{S(0) = 2\} \mid \{S(-t) = 1\}] \\
&= \Pr[\{E(0)\} \mid \{S(0) = 1\}, \{S(-t) = 1\}] \\
&\quad \cdot \Pr[\{S(0) = 1\} \mid \{S(-t) = 1\}] \\
&\quad + \Pr[\{E(0)\} \mid \{S(0) = 2\}, \{S(-t) = 1\}] \\
&\quad \cdot \Pr[\{S(0) = 2\} \mid \{S(t) = 1\}] \\
&= \lambda_2\, dt + (\lambda_1 - \lambda_2)\Pr[\{S(0) = 1] \mid \{S(-t) = 1\}] \\
&= \lambda_2\, dt + (\lambda_1 - \lambda_2)\Pr[\{S(t) = 1] \mid \{S(0) = 1\}] \\
&= \{\lambda_2 + (\lambda_1 - \lambda_2)[\beta/(\alpha + \beta) + \alpha/(\alpha + \beta)e^{-(\alpha+\beta)t}]\}\, dt
\end{aligned} \quad (16)$$

where now t represents time units into the past.

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

$$r(t) = (\lambda_1\beta + \lambda_2\alpha)/(\alpha + \beta) + [\alpha\beta(\lambda_1 - \lambda_2)^2/((\alpha + \beta)(\lambda_1\beta + \lambda_2\alpha))]\, e^{-(\alpha+\beta)t}. \quad (17)$$

Thus the general behavior of the *apparent* contact rate shows an exponential decay into the past, to a steady-state value $\bar{\lambda} = (\lambda_1\beta + \lambda_2\alpha)/$

$(\alpha + \beta)$, 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.

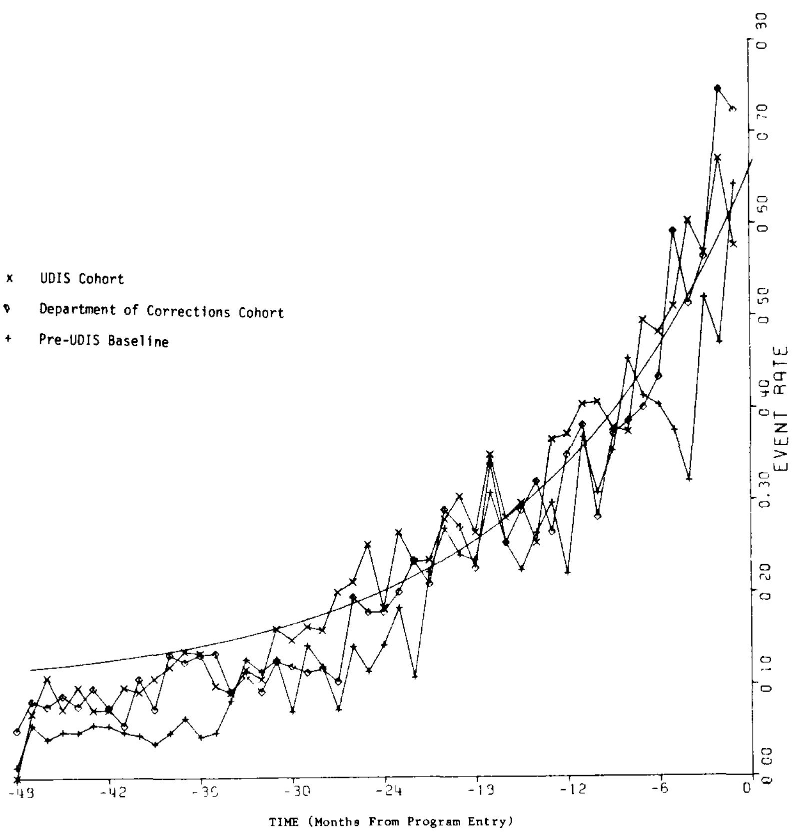

Figure 4. Event rate using a Markov model ($\lambda_1 = 0.8$, $\lambda_2 = 0.02$, $\alpha = 0.063$, $\beta = 0.007$).

Figure 4 shows $r(t)$ from equation (17) for $\lambda_1 = 0.8$, $\lambda_2 = 0.02$, $\alpha = 0.063$ and $\beta = 0.007$, 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.

4. CONCLUSIONS

This paper points out one problem inherent to using data obtained from experimental [1, 2] and quasi-experimental [5] 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.

These 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.

The selection artifact may not be unique to delinquency programs. Automobile drivers are penalized if they accumulate more than a certain number of "points" over a set period of time. The number of points for each infraction is based on the severity of the infraction. License revocation data can be investigated to see if the same "suppression effect" can be found in comparing driving records before and after the penalty is imposed.

We do not mean to imply that the entire "suppression effect" noted by Murray et al. [5] is attributable to an artifact, but rather that some part of it might well be artifactual. It may be possible (with enough effort) to determine the actual selection rules used by judges, then to model these rules to estimate their 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 (a) there is a contact now (at $t = 0$), and (b) there have been at least k ($k \geq 2$) contacts within a time interval of τ. We would then have (see equation (11))

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

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 and on there having been k prior contacts within the time interval τ. To calculate $f(x)$ note that

$$\begin{aligned} F(x) &= P[T_1 \le x \mid \{\text{selection}\}], \quad 0 \le x < \tau \\ &= P[T_1 \le x \mid T_k \le \tau] \\ &= 1 - P[T_1 > x \mid T_k \le \tau] \\ &= 1 - P[T_k \le \tau \mid T_1 > x] P[T_1 > x] / P[T_k \le \tau] \end{aligned}$$

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

and, using (4) we find that

$$f(x) = \begin{cases} \lambda e^{-\lambda x}\, P[T_{k-1} \le \tau - x] / P[T_k \le \tau], & 0 \le x < \tau \\ 0, & \text{elsewhere.} \end{cases} \quad \text{(A-3)}$$

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

ACKNOWLEDGMENTS

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

REFERENCES

1. L. T. Empey and L. L. Erickson. *The Provo Experiment*, Lexington Books, Lexington, Mass., 1972.
2. L. T. Empey and S. G. Lubeck, *The Silverlake Experiment*, Aldine Publishing, Chicago, 1971.
3. R. McCleary, A. Gordon, D. McDowell and M. D. Maltz, "A Reanalysis of 'UDIS: Deinstitutionalizing the Chronic Offender,'" Center for Research in Criminal Justice, University of Illinois at Chicago Circle, November 1978.

4. C. A. MURRAY, Statement before the U.S. Senate Subcommittee on Juvenile Delinquency, Apr. 12, 1978.
5. C. A. MURRAY, D. THOMSON AND C. B. ISRAEL, "UDIS: Deinstitutionalizing the Chronic Juvenile Offender," American Institutes for Research, Washington, D.C., January 1978.
6. *New York Times*: Editorial, June 19, 1978; Letter to the Editor, June 30, 1978.
7. R. M. TERRY, "The Screening of Juvenile Offenders," *J. Criminal Law Criminol. Police Sci.* **58,** 172–181 (1967).