MODELS FOR THE EVALUATION OF TREATMENT-RELEASE CORRECTIONS PROGRAMS

by

EDWARD H. KAPLAN

Technical Report No. 162

Operations Research Center
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

May 1979

Prepared under Grant 78NI-AX-0007 from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

FOREWORD

The Operations Research Center at the Massachusetts Institute of Technology is an interdepartmental activity devoted to graduate education and research in the field of operations research. The work of the Center is supported, in part, by government contracts and grants. This paper was supported by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice under Grant No. 78NI-AX-0007.

Richard C. Larson
Jeremy F. Shapiro

Co-Directors

ABSTRACT

This paper presents a discussion of methods available for conducting model-based evaluations of treatment-release corrections programs. A general model of rearrest patterns over time is described along with a numerical example illustrating model behavior under alternative assumptions. Classical and Bayesian methods for the estimation of model parameters are reviewed, as are complementary model-based evaluation procedures.

ACKNOWLEDGEMENTS

I'd like to thank Dick Larson for his blackboard antics, Keith Stevenson and Vicki Bier for the time they spent discussing this subject with me, and Arnie Barnett for his help and his humour.

MODELS FOR THE EVALUATION OF TREATMENT - RELEASE CORRECTIONS PROGRAMS

I. INTRODUCTION

Within the corrections component of the criminal justice system, a range of programs aimed at the rehabilitation of selected individuals has been established. These programs (including prison, parole, residential centers for drug offenders, alcohol abuse counselling, etc.) all share the following feature in common: individuals committed to a program are subjected to a period of "treatment"; upon satisfactory completion of the treatment period, these individuals are "released" (hence the term "treatment-release program"). Of interest to the officials of such programs is the event that a randomly chosen program client commits an offense after release; the likelihood of this event is termed the "recidivism probability."

While the concept of a recidivism probability poses no immediate difficulty, the measurement of recidivism is not an easy task. In a clever paper, Blumstein and Larson (1969) discussed measurement problems which arise from alternative definitions of recidivism, and from improper interpretation of sample statistics. If we allow recidivism to refer solely to the event where a program client commits an offense after release, then recidivism cannot be measured directly.

When an individual commits a crime, there is no guarantee that (s)he will be apprehended. Of all individuals who commit crimes, some fraction will in fact be arrested by the police. Placed into the context of a treatment-release corrections program, only those recidivists who are rearrested are in fact observed as having recidivated; indeed, the difference between rearrest levels and true recidivism levels may be greater than one might expect (Barnett and Stabile, 1979). Whether or not an offender is apprehended depends upon police performance as well as upon the nature of the offense. Thus, the methods of this paper will be presented as applied to rearrest patterns over time, as this is the type of data which is frequently available.

What is found in the remainder of this paper is a discussion of methodology for conducting model-based evaluations of treatment-release corrections programs. Model-based techniques have proved useful in evaluating police patrol programs (Larson, 1975; Kaplan, 1978a), and it is felt that the advantages provided by the modeling approach can carry over to the corrections area. We begin with the description of a general model for rearrest patterns over time; the behavior of this model is examined under alternative assumptions in a numerical example. Classical and Bayesian estimation methods are presented, followed by a discussion of model-based evaluation procedures. The paper concludes with a brief discussion of possible extensions to the work reported here.

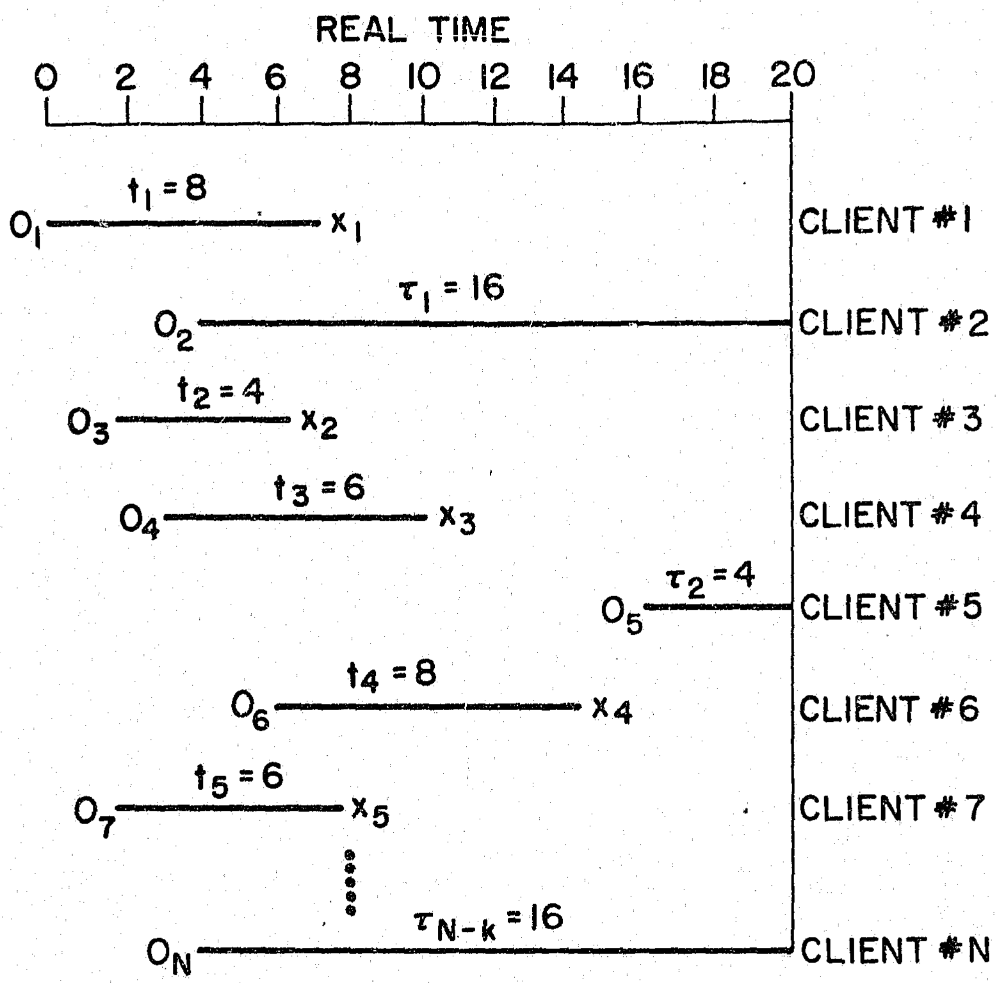
II. A GENERAL MODEL OF REARREST PATTERNS

Consider any corrections program which subjects its clients to a treatment period after which they are released. Ideally, clients released from treatment return to the community as law - abiding citizens. Realistically, sizeable fractions of program populations are known to recidivate. Of those who do recidivate, some are apprehended and rearrested.

In the model to be presented, we exploit the similarities between the rearrest process and a branch of probabilistic reasoning known as reliability theory (see Chapter 4 in Tsokos (1972), Chapter 13 in Hillier and Lieberman (1974).[1] Figure 1 depicts our observation of a corrections program which releases clients at different points in time; in total, N clients are released. Of these N individuals, some are rearrested, while the others are not rearrested during the period of time allowed for observation.

For all clients in the program, our model will measure time according to time from individual release, hence all of our arguments are conditioned on release occurring at time zero. An observed rearrest is referred to as a "failure", and the length of time that elapses between an individual's release and failure is denoted as the "time until failure". Hence, the statement "Five failures occurred by the eighth day after release" is interpreted to mean that five (of N) individuals were rearrested within eight days of the particular day on which they were each individually released.

Consider an individual who is released from treatment at time $0 (t_R = 0)$. We are interested in the probability that this same individual will be rearrested at some future time t given release at $t_R = 0$. Let



O_n = TIME OF n th RELEASE $n = 1, 2, \ldots, N$
x_i = TIME OF i th REARREST $i = 1, 2, \ldots, k$
t_i = TIME FROM RELEASE UNTIL i th FAILURE $i = 1, 2, \ldots, k$
τ_j = OBSERVED TIME FROM RELEASE OF j th SUCCESS, $j = 1, 2, \ldots, N-k$

Figure 1

A Rearrest Pattern Over Time Measured According To Time From Release this <u>probability of failure</u> be denoted by $p_F(t)$. In order to derive expressions for $p_F(t)$ and other performance measures, we need to make a few simplifying assumptions.

(i) All individuals fail independently of each other.

(ii) Pr {any client fails in $(t,t + dt)$ | ultimate failure, but not in $(0,t)$, $t_R = 0$}
$= \phi(t|F)dt$.

(iii) Pr {any client does not fail in $(t,t + dt)$ | ultimate failure, but not in $(0,t)$, $t_R = 0$}
$= 1 - \phi(t|F)dt$.

(iv) The fraction of the population that will ultimately fail is given by r; $0 \leq r \leq 1$.

From reliability theory, it is well known that these assumptions determine the conditional probability of failure by time t given $t_R = 0$ and ultimate failure to be (Hillier and Lieberman, (1974))

$$p_F(t|F) = 1 - e^{-\int_0^t \phi(x|F)dx}, \qquad t > 0 \qquad (1)$$

subject to:

(i) $\phi(t|F) \geq 0 \ \forall \ t > 0$.

(ii) $\int_0^\infty \phi(t|F)dt = \infty$

The unconditional probability of failure by time t after release, $p_F(t)$, is then given by

$$p_F(t) = r(1 - e^{-\int_0^t \phi(x|F)dx}). \qquad t > 0 \qquad (2)$$

$$0 \leq r \leq 1$$

It is apparent that the behavior of this model is completely dependent on the nature of the function $\phi(t|F)$. This function, which is referred to as the hazard function, dictates the probability of failure in the next time instant for those clients who have yet to fail but will ultimately fail. Several functions come to mind for $\phi(t|F)$. These are shown in Figure 2. Curves of Type I are of the form $\phi(t|F) = \lambda$; such an assumption implies that $p_F(t|F)$ will take on the simple negative exponential distribution. Stollmack and Harris (1974) studied this model, they also assumed that the fraction of utlimate failures was equal to one, a rather restrictive assumption. Maltz and McCleary (1977) studied this model without the $r = 1$ assumption. Both of these models will be examined later on in this paper. Type II and Type III curves involve increasing or decreasing propensities to be rearrested over time. Some of these curves can be formulated as $\phi(t|F) = \alpha\beta t^{\beta-1}$, which implies that $p_F(t|F)$ takes on the Weibull distribution (Freund, 1971:117); a model of this sort is illustrated later on. A more complicated model involving an exponential - type hazard function is discussed by Bloom (1978). In the appendix, we show that Bloom's model performs equivalently to the Maltz-McCleary model. Hence, Bloom's model will not be reviewed. Of course, plausible arguments for more complicated curves such as Types IV and V can be made; these would lead to still more complex forms for $p_F(t|F)$. The choice of an appropriate function for $\phi(t|F)$ is a data analysis question not pursued in this paper; since there is no one correct function $\phi(t|F)$ which works for all situations, the results which follow will be notated for the general case.

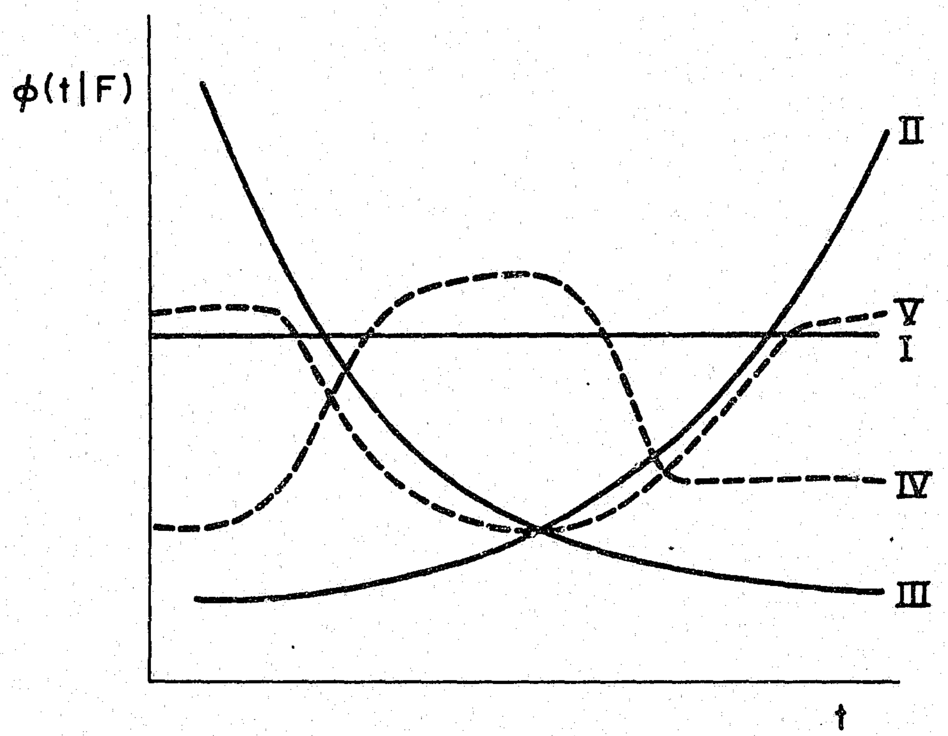


Figure 2

Possible Curves for $\phi(t|F)$

Performance Measures for the General Model

(i) Ultimate Probability of Failure: r

The ultimate probability of failure, $r = p_F(\infty)$, is directly estimated along with whatever parameters accompany the hazard function $\phi(t|F)$; this estimation problem is discussed in Section IV.

(ii) Expected Time Until Failure: $\bar{t}_F$

We have already computed the unconditional probability of failure by time t to be

$$p_F(t) = r(1 - e^{-\int_0^t \phi(x|F)dx}). \qquad t > 0$$

Differentiating this expression with respect to time from release provides us with the pseudo - pdf[2] for clients' individual times until failure

$$f_{t_F}(t) = r\phi(t|F)e^{-\int_0^t \phi(x|F)dx}. \qquad t > 0 \qquad (3)$$

The expected time until failures for ultimate failures is thus formulated as

$$\bar{t}_F = \frac{1}{r}\int_0^\infty t\ f_{t_F}(t)dt. \qquad (4)$$

Note that unless r = 1, the unconditional expected time until failure is infinite, since

E(time to failure) = E(time to failure|ultimate failure)·Pr {ultimate failure}

+ E(time to failure|ultimate success)·Pr {ultimate success}

$$= \bar{t}_F \cdot r + \infty \cdot (1-r)$$

$$= \infty$$

(iii) Median Time Until Failure: $t_{F(.5)}$

The median time until failure for ultimate failures is given by the equation

$$t_{F(.5)} = p_F^{-1}(.5r). \qquad (5)$$

This is the time by which 50% of all eventual rearrests will have occurred.

(iv) Safety Time: $t^*(\varepsilon)$

The performance measure $t^*(\varepsilon)$ satisfies the equation

$$\Pr\{\text{individual fails in } (t^*,\infty) \mid \text{didn't fail in } (0,t^*) \text{ and } t_R=0\} = \varepsilon;$$

and hence defines the safety time at risk level ε (see Bloom (1978)). Intuitively, if one wishes to observe a client after release until the rearrest probability of that client is less than or equal to ε, then one must observe that client for at least $t^*(\varepsilon)$ time units after release at $t_R = 0$. The safety time is found by solving

$$t^*(\varepsilon) = p_F^{-1}\left(\frac{r - \varepsilon}{1 - \varepsilon}\right), \quad 0 < \varepsilon < r. \qquad (6)$$

(v) Probability Mass Function of the Number of Rearrests: $P_n(t)$

Recall that by assumption, all individuals fail independently of one another. Since the probability of failure by time t given release at $t_R = 0$, $p_F(t)$, applies to any client, the probability that exactly n individuals have failed by time t is given by the well known binomial pmf

$$P_n(t) = \binom{N}{n} [p_F(t)]^n [1 - p_F(t)]^{N-n} \tag{7}$$

$$n = 0,1,..,N$$

$$t \geq 0 .$$

Note that the probabilities

$$P_n(\infty) = \binom{N}{n} r^n (1 - r)^{N-n}$$

$$n = 0,1,..,N \tag{8}$$

$$t \geq 0$$

may be interpreted as long run probabilities of failure, in that equation (8) determines the probability distribution of the ultimate failure population.

(vi) Expected Number of Failures: $\bar{n}(t)$

The expected number of failures that have occurred by time t is simply

$$\bar{n}(t) = Np_F(t). \tag{9}$$

since n(t) is a binomially distributed random variable.

(vii) Variance of the Number of Failures: $\sigma_n^2(t)$

This measure is also easily obtained; it is given by

$$\sigma_n^2(t) \doteq Np_F(t)\ (1 - p_F(t)). \qquad (10)$$

It is interesting to note that since $0 \leq p_F(t) \leq 1$, $\sigma_n^2(t)$ will achieve a maximum when $p_F(t) = 1/2$. This implies that regardless of the functional form of $p_F(t)$, the maximum achievable variance of this model is equal to N/4. Under the conditions of this model, at worst one can be 95% certain that the observed number of failures by any time t is within $\pm\sqrt{N}$ of $\bar{n}(t)$.

(viii) Gausian Approximation to $P_n(t)$: $f_{n(t)}(x)$

Most correctional programs involve a large number of clients. In such situations where N is large, the calculations of $P_n(t)$ are both tedious and perhaps unrewarding. However, if N is sufficiently large (i.e., if $Np_F(t)$ and $N(1 - p_F(t))$, are both greater than 5 (Freund, 1971:177)), one may approximate the discrete binomial distribution of n(t) by a Gaussian distribution with mean $\bar{n}(t)$ and variance $\sigma_n^2(t)$. If we allow the interval $(n - 1/2, n + 1/2]$ to represent the integer number of failures n, then the number of failures which occur by a given time t may be approximated as a continuous random variable with pdf

$$f_{n(t)}(x) = \frac{1}{\sigma_n(t)\sqrt{2\pi}} e^{-1/2\left[\frac{x - \bar{n}(t)}{\sigma_n(t)}\right]^2}. \qquad (11)$$

$$-\infty < x < \infty$$

The corresponding steady state pdf is obtained by setting $t = \infty$ in the above formulation. For large values of N, these distributions $f_{n(t)}(x)$ will be quite accurate.

(ix) <u>Time by Which the Probability of Failure Equals k/N: T_k</u>

One may also be interested in the time until the k^{th} failure. While there is no theoretical problem formulating the expected time until the k^{th} failure given that at least k failures ultimately occur, the computations involved are prohibitively difficult. To gain some indication of the timing of rearrests, it seems reasonable to examine

$$T_k \equiv \bar{n}^{-1}(k). \qquad (12)$$

This statistic reports the time by which the probability of failure equals k/N; we will use T_k in Section V to compute <u>fractile</u> times until failure for ultimate failures.

In summary, this section has presented a general model which can provide a framework for analyzing rearrest patterns over time. Having discussed this model, it is useful to examine the differences in model performance that result from the choice of alternative hazard functions. Such an example is presented in the next section.

III. AN ILLUSTRATIVE EXAMPLE

Consider a treatment-release corrections program which serves a client population of N = 100. Suppose that a preliminary review of rearrest data has revealed that 25 of the program's 100 clients were rearrested within 6 months after release (hence $p_F(6)$ is estimated to equal .25). Program officials are committed to evaluate the program after 24 months of exposure data have been collected. The program will be considered a success if $p_F(24) < .40$. In the meantime, the program staff would like some indications of the range of rearrest patterns that could occur over time under alternative assumptions governing the rearrest process.

Three conjectures are of particular interest to the program staff; each may be formulated as a model consisted with the data point $p_F(6) = .25$.

Model 1 (Stollmack-Harris)

Let $\phi(t|F) = \lambda$, $\lambda > 0$, and assume $r = 1.0$.

Model 2 (Maltz-McCleary)

Let $\phi(t|F) = a$, $a > 0$, and assume $r = 0.5$.

Model 3 (Weibull)

Let $\phi(t|F) = \alpha\ \beta\ t^{\beta-1}$, α, $\beta > 0$, and assume $r = 0.5$.

The implications of these postulates may be examined in some detail. Table I presents the formulas used to compute the measures associated with the models, while Table II reports numerical values for selected measures.

If we direct our attention to Figure 3, we notice that the three models do represent quite different rearrest patterns over time.

TABLE I

FORMULAS FOR PERFORMANCE MEASURES

PERFORMANCE MEASURES	EQUATION	STOLLMACK-HARRIS	MALTZ-MCCLEARY	WEIBULL
$p_F(t)$	(2)	$1 - e^{-\lambda t}$	$r(1 - e^{-at})$	$r(1 - e^{-\alpha t^{\beta}})$
$\bar{t}_F$	(4)	$\frac{1}{\lambda}$	$\frac{1}{a}$	$(\frac{1}{\alpha})^{\frac{1}{\beta}}\Gamma(1 + \frac{1}{\beta})$
$t_F(.5)$	(5)	$-\frac{1}{\lambda}\ln(.5)$	$-\frac{1}{a}\ln(.5)$	$\left[-\frac{1}{\alpha}\ln(.5)\right]^{\frac{1}{\beta}}$
$t^*(\varepsilon)$	(6)	N.A.	$-\frac{1}{a}\ln(\frac{\varepsilon(1-r)}{r(1-\varepsilon)})$	$\left[-\frac{1}{\alpha}\ln(\frac{\varepsilon(1-r)}{r(1-\varepsilon)})\right]^{\frac{1}{\beta}}$
$\bar{n}(t)$	(9)	$N(1 - e^{-\lambda t})$	$Nr(1 - e^{-at})$	$Nr(1 - e^{-\alpha t^{\beta}})$
$\sigma_n^2(t)$	(10)	$N(1 - e^{-\lambda t})e^{-\lambda t}$	$Nr(1 - e^{-at}) \cdot (1-r + re^{-at})$	$Nr(1 - e^{-\alpha t^{\beta}}) \cdot (1-r + re^{-\alpha t^{\beta}})$
T_k	(12)	$-\frac{1}{\lambda}\ln(1 - \frac{k}{N})$	$-\frac{1}{a}\ln(1 - \frac{k}{Nr})$	$\left[-\frac{1}{\alpha}\ln(1 - \frac{k}{Nr})\right]^{\frac{1}{\beta}}$

TABLE II

NUMERICAL RESULTS

PERFORMANCE MEASURE		STOLLMACK-HARRIS (λ = .05)	MALTZ-MCCLEARY (a = .12)	WEIBULL (α = .28, β = .50)
r		1.0	0.5	0.5
$t_{F(.5)}$	(months)	13.9	6.0	6.0
$\bar{t}_F$		20.0	8.3	25.5
$t^*(.1)$	(months)	N.A.	18.3	61.6
$t^*(.2)$		N.A.	11.6	24.5
$t^*(.3)$		N.A.	7.1	9.2
$t^*(.4)$		N.A.	3.4	2.1
T_1	(months)	2.1	0.2	0.0
T_{20}		4.5	4.3	3.3
T_{40}		10.2	13.4	33.0

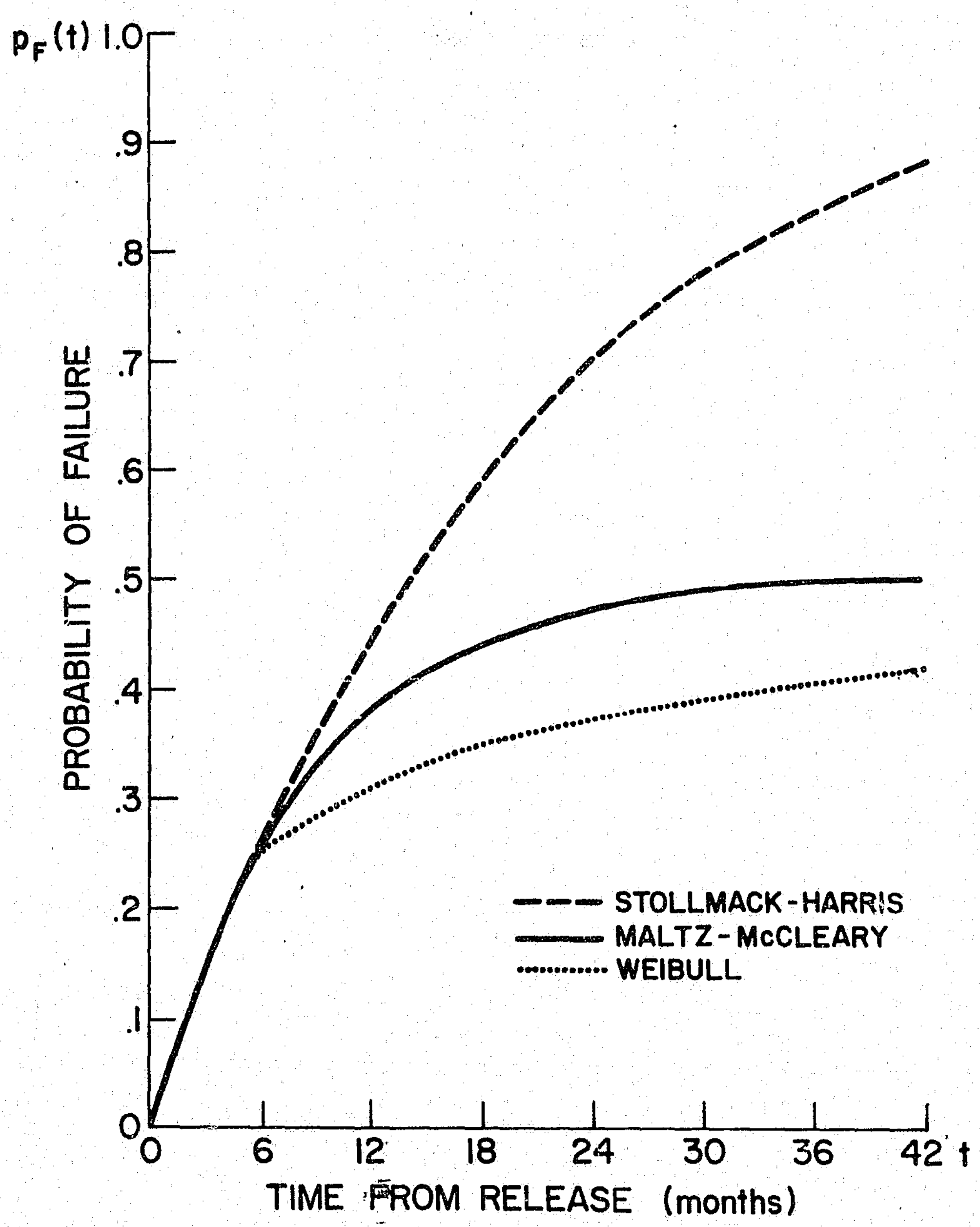


Figure 3

Failure Probability Over Time

Recalling that an evaluation is to be performed on the basis of 24 months of exposure data, the model-based values of $p_F(24)$ are useful indices for shaping prior expectations of program performance. As an example, our Weibull model demonstrates that even if $p_F(24) = .37 < .40$, the criterion set a priori by program administrators, the long run probability of re-arrest can equal .50. Thus, program success in the short run is not inconsistent with long run program failure; evaluations of such programs should take this possibility into account.

If we now consider the timing of rearrests, some differences in model behavior are noteworthy. Although the Maltz-McCleary and Weibull models produce equivalent median times until failure, the Weibull expected time until failure is more than three times that of the Maltz-McCleary model. Of the three models considered, the Maltz-McCleary model clearly exhibits the most rapid failure process over time for ultimate failures; this is best reflected by Figure 3.

Figure 4 presents a graph of the safety time $t^*(\varepsilon)$ versus ε for the Maltz-McCleary and Weibull models. The Weibull model is clearly conservative in its implications. Only after an arrest-free release of 61.6 months can one be 90% certain that a client will not fail according to the Weibull model. At an equivalent 90% confidence level, the Maltz-McCleary model requires 18.3 months of arrest free releases.

To examine the uncertainty associated with these models, the variance of the number of failures is plotted as a function of time in Figure 5 for each of our three sets of assumptions. All three models reach the maximum achievable variance of N/4, since $p_F(t)$ approaches 1/2 as t approaches infinity for the Maltz-McCleary and Weibull models,

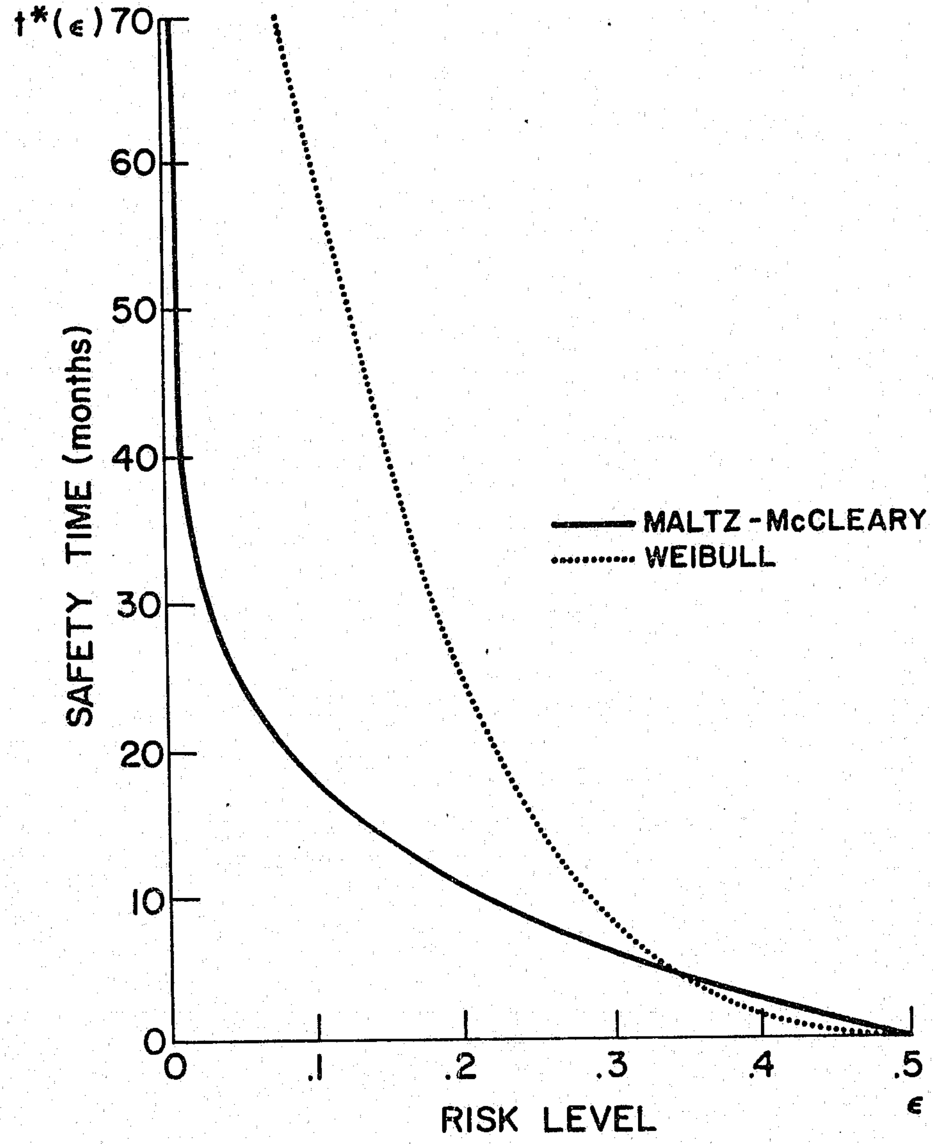


Figure 4

Safety Time Versus Risk Level

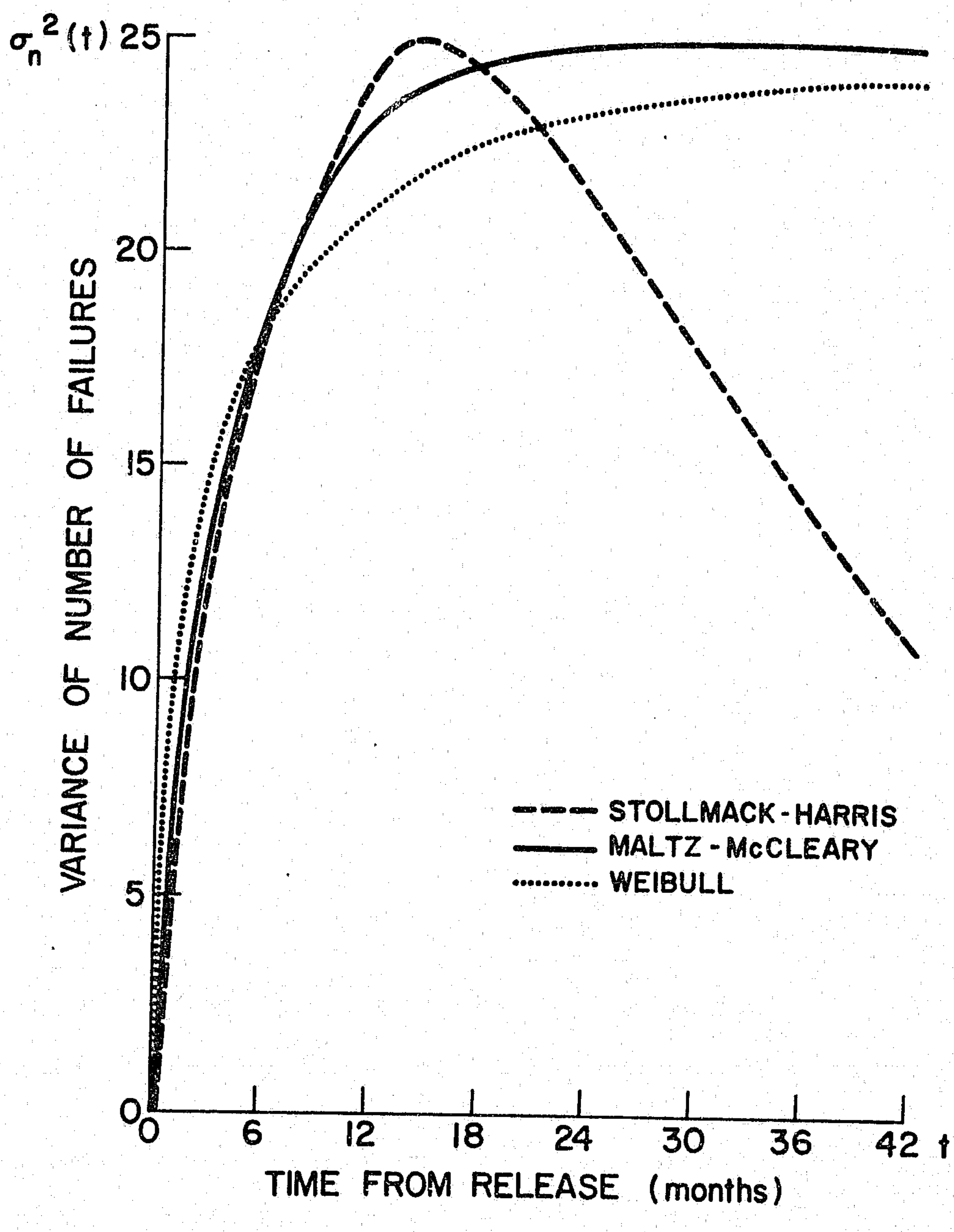


Figure 5

Variance of Failure Population Over Time while $p_F(13.9) = 1/2$ for the Stollmack-Harris model.

Finally, Figures 6, 7 and 8 present the Gaussian approximations to $P_n(t)$ for $t = 6$, 24 and ∞. At $t = 6$, all three models are identical. The distributions are quite distinct at $t = 24$, with the Stollmack-Harris model translated the furthest to the right, and the Weibull model the furthest to the left. As t approaches infinity, the Stollmack-Harris model produces an infinite spike at $n(\infty) = N$. Of course, the Maltz-McCleary and Weibull models reproduce each other for this case.

It is apparent that the behavior of the models can be drastically different at various points in time. This stems from the alternative formulations of $p_F(t)$ which in turn depend upon the assumptions governing the behavior of $\phi(t|F)$, the conditional hazard function. While the models demonstrated here do not by any means exhaust the world of possible models, they do illustrate the different types of model behavior achievable via the specification of alternative hazard functions.

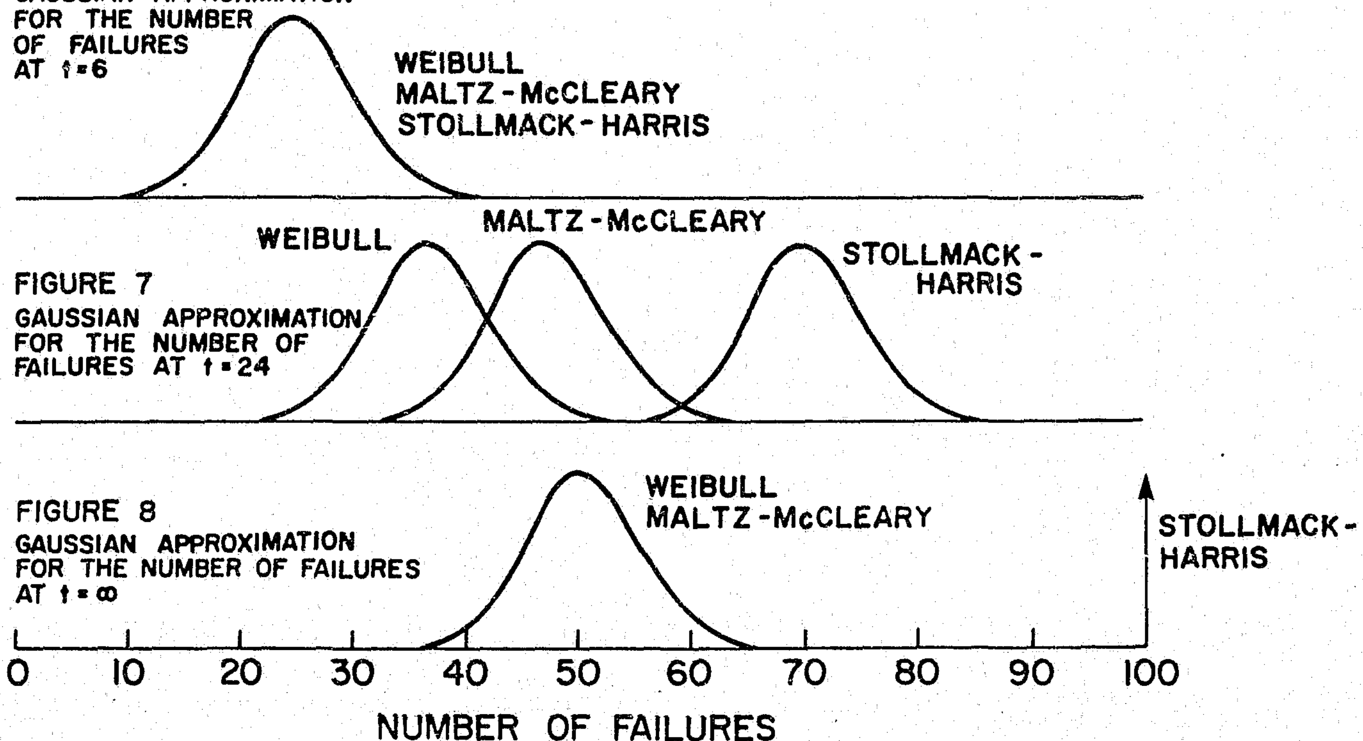


FIGURE 6
GAUSSIAN APPROXIMATION FOR THE NUMBER OF FAILURES AT $t = 6$

FIGURE 7
GAUSSIAN APPROXIMATION FOR THE NUMBER OF FAILURES AT $t = 24$

FIGURE 8
GAUSSIAN APPROXIMATION FOR THE NUMBER OF FAILURES AT $t = \infty$

IV. ESTIMATION OF PARAMETERS FOR THE REARREST MODEL

In order to utilize in practice the general formulation developed thus far, we need to consider some reasonable techniques for estimating r and the parameters of $\phi(t|F)$. To this end, both maximum likelihood and Bayesian methods are suggested. The data aggregation scheme presented in conjunction with this discussion is attributable to Stollmack and Harris (1974).

Recall our discussion of Figure 1 from Section II. Suppose that of the N individuals who were released, k have been rearrested by the time we begin our analysis. If we let t_i correspond to the time from release until the i^{th} failure $(i = 1,2,..,k)$, then the likelihood of observing these k failures at the times they occurred under the conditions of our model is given by

$$L\{k, \underset{\sim}{t}\} = \prod_{i=1}^{k} r\,\phi(t_i|F)e^{-\int_0^{t_i}\phi(x|F)dx} \quad . \tag{13}$$

Similarly, let τ_j represent the time from release that the j^{th} client $(j = 1,2,..,N-k)$ has been observed to remain unarrested. The probability of observing this combination of the N-k success times is given by

$$L\{N-k, \underset{\sim}{\tau}\} = \prod_{j=1}^{N-k}(1 - r + re^{-\int_0^{\tau_j}\phi(x|F)\,dx}). \tag{14}$$

(i) Maximum Likelihood Estimation

Let $\phi(t;\underset{\sim}{\psi}|F)$ denote the conditional hazard function where $\underset{\sim}{\psi}$ is the set of parameters contained in this function (e.g. for the Weibull model, $\underset{\sim}{\psi} = \{\alpha, \beta\}$). The overall likelihood of observing a particular pattern of k failures over time according to our model is given by

$$L\{k, N, \underset{\sim}{t}, \underset{\sim}{\tau} | r, \underset{\sim}{\psi}\} = \left[\prod_{i=1}^{k} r\phi(t_i; \underset{\sim}{\psi} | F) e^{-\int_0^{t_i} \phi(x; \underset{\sim}{\psi} | F) dx} \right] \cdot \left[\prod_{j=1}^{N-k} (1 - r + r e^{-\int_0^{\tau_j} \phi(x; \underset{\sim}{\psi} | F) dx}) \right]. \qquad (15)$$

To find the maximum likelihood estimates for r and $\underset{\sim}{\psi}$, one must solve the optimization problem

$$\max_{r, \underset{\sim}{\psi}} L\{k, N, \underset{\sim}{t}, \underset{\sim}{\tau} | r, \underset{\sim}{\psi}\} \qquad (16)$$

subject to $0 < r < 1$, constraints on $\underset{\sim}{\psi}$

In general, this is not an easy problem to solve. Maximum likelihood estimates have been obtained analytically for the Stollmack-Harris model (Stollmack-Harris, 1974), numerically for Bloom's model (Bloom, 1978) and numerically for the Maltz-McCleary model under the special condition that $\tau_j = \tau$, $j = 1, 2, .., N-k$ (Maltz and McCleary, 1977). To obtain maximum likelihood estimates for more complicated forms of $\phi(t|F)$ may require the use of non-linear programming routines.

(ii) Bayesian Estimation

In Bayesian analysis, we allow both subjective and objective information to play a role in our model (Freund, 1971: 280-281). The parameters of interest, r and $\underset{\sim}{\psi}$, are not viewed as being fixed and unchanging as is the case with classical techniques such as maximum likelihood estimation. Rather, r and $\underset{\sim}{\psi}$ are assumed to behave as random variables with *a priori* probability distributions. These distributions may be objectively or subjectively derived.

As an example, assume for the moment that r and ψ are independent within the Bayesian scheme of things. A corrections program official might make a series of statements of the form:

"The probability that some fraction not exceeding r of our clients will fail is given by F(r)."

These statements may be based on both the past experience of other programs and personal convictions regarding the likelihood of program success. This subjective cumulative distribution F(r) is then converted to a density function by examining successive differences (e.g., F(.1) - F(.05), F(.15) - F(.1), ... F(1.0) - F(.95)) and fitting a curve to the resultant histogram. Such a prior distribution f(r) is interpreted as the probability distribution of ultimate failure likelihoods across the population of programs similar to the one in question. The true value of r that will actually be observed is treated as a random selection from this population.

Suppose that a program administrator specified the following values for r and F(r):

r	F(r)
.2	.10
.4	.50
.6	.80
.8	.95
1.0	1.00

The prior distribution graphed in Figure 9 would result. This distribution formally represents prior expectations of ultimate rearrest probabilities.

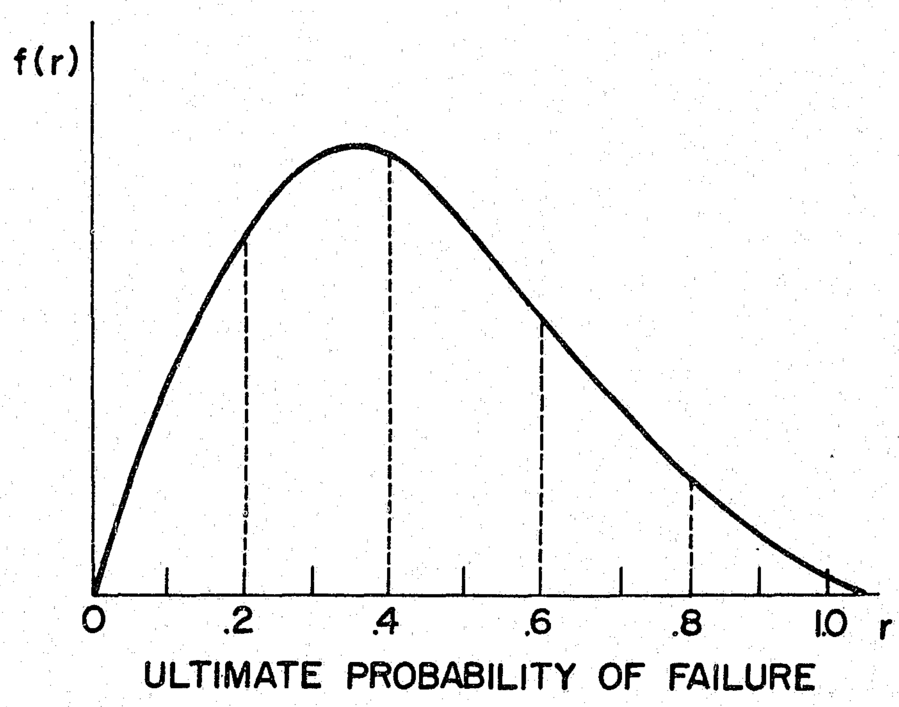


Figure 9

Prior Distribution of the Ultimate Probability of Failure

Let $f(r, \underset{\sim}{\psi})$ represent the joint prior distribution of r and $\underset{\sim}{\psi}$. Since r and $\underset{\sim}{\psi}$ are now being viewed as random variables, the likelihood function of equation (15) is nothing more than the conditional probability of having observed a particular rearrest pattern given specific values of r and $\underset{\sim}{\psi}$. What we wish to compute is the joint conditional distribution of r and $\underset{\sim}{\psi}$ having observed a particular rearrest pattern characterized by k failures out of N released clients, and times to failure $\underset{\sim}{t}$, observed success times $\underset{\sim}{\tau}$.

This posterior distribution of r and $\underset{\sim}{\psi}$ is formulated via the use of Bayes' Rule (Freund, 1971: 280-281)

$$g(r, \underset{\sim}{\psi}|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) = \frac{f(r, \underset{\sim}{\psi})L\{k, N, \underset{\sim}{t}, \underset{\sim}{\tau}|r, \underset{\sim}{\psi}\}}{R(k, N, \underset{\sim}{t}, \underset{\sim}{\tau})} \tag{17}$$

where $R(k, N, \underset{\sim}{t}, \underset{\sim}{\tau})$ is a normalizing constant. The posterior distribution of r may be found by integrating out over $\underset{\sim}{\psi}$

$$h(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) = \int_{\psi_1 \varepsilon \underset{\sim}{\psi}} \cdots\cdots \int_{\psi_n \varepsilon \underset{\sim}{\psi}} g(r, \underset{\sim}{\psi}|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})d\psi_1 \ldots d\psi_n . \tag{18}$$

The posterior expected value of r is then found by computing

$$E(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) = \int_{r=0}^{1} r\, h(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})dr. \tag{19}$$

Similarly, the posterior expected values of each of the performance measures discussed in Section II can be found. Let a generic performance measure be denoted by $M(r, \underset{\sim}{\psi})$; the posterior expected value of this measure is given by

$$E[M(r, \underset{\sim}{\psi}) | k, N, \underset{\sim}{t}, \underset{\sim}{\tau}] = \int_{r=0}^{1} \int_{\psi_1 \varepsilon \underset{\sim}{\psi}} \cdots \int_{\psi_n \varepsilon \underset{\sim}{\psi}} M(r,\psi) g(r,\underset{\sim}{\psi} | k,N,\underset{\sim}{t},\underset{\sim}{\tau}) d\psi_1 \ldots d\psi_n \, dr. \quad (20)$$

Now, since the calculations in the equations of this Bayesian analysis require nothing more complex than integration, the Bayesian estimates discussed can be obtained numerically. It would be possible to write a computer program to perform these calculations for any given function $\phi(t|F)$, though the design of such a program has not yet been attempted.

V. EVALUATION ISSUES

In the introduction to this paper, we stressed the difference that exists between observed rearrest rates and true recidivism rates. Since we have not attempted to account for the relative influence the police have on the rearrest process, the models which have been discussed should be used in conjunction with controlled evaluation designs (see Campbell and Stanley, 1966) if the comparison of two corrections programs is being pursued.

The manner in which our model is used for evaluation purposes depends upon whether the estimation approach chosen is classical or Bayesian. The differences resulting from these alternative approaches are illustrated throughout.

The first measure of evaluative interest is the ultimate rearrest probability r. In general, program success is seen to vary inversely with the value of r.[3] It may be established a priori by program officials that one program target is the achievement of an r value less than some desired tolerance level r^*. If r was estimated via maximum likelihood techniques, then for large N, the null hypothesis H_o: $r = r^*$ may be tested using

$$Z = \frac{\hat{r} - r^*}{\sqrt{\frac{\hat{r}(1-\hat{r})}{N}}} \tag{21}$$

where $\hat{r}$ is the maximum likelihood estimate of r, and Z is distributed as a standardized Gaussian random variable (Freund, 1971:329).[4] The rationale for this test stems from the Gaussian approximation $f_{n(t)}(x)$ discussed in Section II.

The Bayesian version of this test would consist of computing the probability that r is truly less than r^*. This is computed as

$$\Pr\{r < r^* | k, N, \underset{\sim}{t}, \underset{\sim}{\tau}\} = \int_0^{r^*} h(r | k, N, \underset{\sim}{t}, \underset{\sim}{\tau})\, dr \qquad (22)$$

where $h(r | k, N, \underset{\sim}{t}, \underset{\sim}{\tau})$ is as defined in Section IV. What constitutes an acceptable likelihood of program success in this instance is a decision problem for program officials.

One would also be interested in the average rearrest probability. To this end, $E(r | k, N, \underset{\sim}{t}, \underset{\sim}{\tau})$ may be found through use of equation (19). A computed value of $E(r | k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) < r^*$ is indicative of program success.

The procedure just presented may also prove useful as indicators of whether or not additional data collection is necessary during the life of a program. Suppose that after some initial fixed period of data collection, k failures out of N releases have occurred. For the classical procedure, compute the maximum likelihood estimate $\hat{r}$, and substitute k/n for r^* in equation (21). If there appears to be no significant difference between $\hat{r}$ and k/N, then perhaps there is no need to continue collecting data, and resources available for this segment of the evaluation may be channeled to other evaluation tasks (e.g. interviews with program clients). The Bayesian analogy consists of substituting k/N for r^* in equation (22); if $\Pr\{r < k/N | k, N, \underset{\sim}{t}, \underset{\sim}{\tau}\}$ is relatively large, then this may also be an appropriate signal to end data collection.

Conversely, if there is a strong disagreement between the observed fraction of failures and the estimated ultimate fraction of failures, maybe more data should be collected, even if the time by which k rearrests have occurred corresponds to the scheduled completion date for the data collection effort. For an example of this sort, data extracted from Pre-Trial Intervention: A Program Evaluation of Nine Manpower-based Pre-Trial Intervention Projects revealed that after one year of release time, 18.3% of all released clients had been rearrested, yet application of the Maltz-McCleary model to this same data yielded a maximum likelihood estimate of $\hat{r} = .43$ (Kaplan, 1978b:23). Since $\hat{r}$ is almost 2 1/2 times as large as k/N in this instance, it might have been a good idea to sustain the data collection effort for this evaluation; if models like those presented here had been available to these evaluators, this finding could have been discovered during the data collection phase.

If we now consider the case where two programs are being compared in a controlled environment, a number of our model-based performance measures may be utilized. Again focusing our attention on the ultimate probability of rearrest r, the null hypothesis H_o: $r_1 = r_2$ may be tested using

$$Z = \frac{\hat{r}_1 - \hat{r}_2}{\sqrt{\frac{\hat{r}_1(1 - \hat{r}_1)}{N_1} + \frac{\hat{r}_2(1 - \hat{r}_2)}{N_2}}} \qquad (23)$$

where:

N_1, N_2 = client populations of the programs;

$\hat{r}_1$, $\hat{r}_2$ = maximum likelihood estimates of the ultimate failure probabilities;

Z = a standardized Gausian random variable.

As with all of the procedures we have discussed which are dependent upon the Gausian approximation $f_{n(t)}(x)$, this test is valid only if N_1 and N_2 are large (Freund, 1971: 332).

The comparison of two programs from a Bayesian perspective using r commends a more graphic analysis. Essentially, the posterior distributions of r for each program may be plotted on the same figure; such a presentation provides a visual method for comparing program performance. An example of such a plot is shown in Figure 10.

The Bayesian approach does allow for numerical comparisons as well. The simplest of such comparisons would be to compute the expected posterior probabilities of ultimate rearrest using equation (19), and check to see which program produced the lower value. A more meaningful comparison involves finding the likelihood that one program produced a lower probability of rearrest than the other program. If we let

$$\underset{\sim}{A}_1 \equiv \{k_1, N_1, \underset{\sim}{t}_1, \underset{\sim}{\tau}_1\}$$

$$\underset{\sim}{A}_2 \equiv \{k_2, N_2, \underset{\sim}{t}_2, \underset{\sim}{\tau}_2\}$$

where the subscripts denote program one and two, then the expected probability that $r_1 < r_2$ is computed as

$$\Pr\{r_1 < r_2 | \underset{\sim}{A}_1, \underset{\sim}{A}_2\} = \int_{r_2=0}^{1} \int_{r_1=0}^{r_2} h_1(r_1|\underset{\sim}{A}_1) h_2(r_2|\underset{\sim}{A}_2) dr_1 dr_2 \qquad (24)$$

where $h(r|\underset{\sim}{A})$ is as defined in equation (18). Conversely, $\Pr\{r_1 > r_2|\underset{\sim}{A}_1, \underset{\sim}{A}_2\} = 1 - \Pr\{r_1 < r_2|\underset{\sim}{A}_1, \underset{\sim}{A}_2\}$. If the result of equation (24) is greater than 1/2, then it would appear that program one has outperformed program two using r as a performance measure.

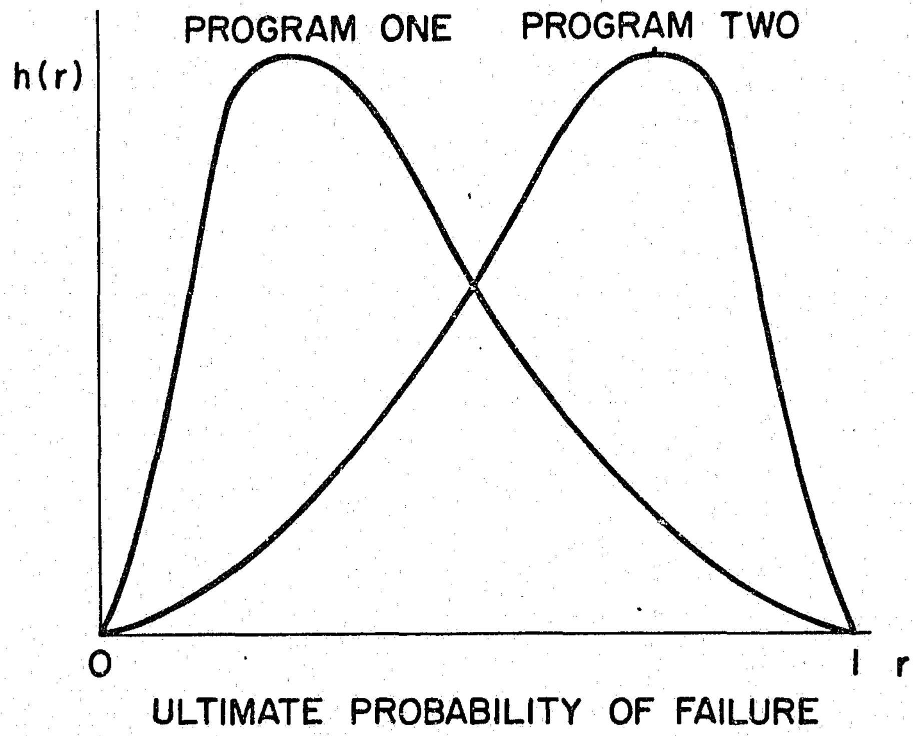


Figure 10

Comparing the Posterior Distributions
Of Ultimate Failure Probabilities

When comparing two programs, the timing of rearrests becomes important. For example, if two programs produced equivalent values for r, one could argue that the program with the lesser value for $\bar{t}_F$ (or $E(\bar{t}_F)$ if you're a Bayesian) was the more successful since such a program quickly distinguishes ultimate failures from the rest of the client population. Indeed, Maltz and McCleary recognized this possibility when they wrote that "Knowledge of a program's failure rate can help in matching programs to participants" (Maltz and McCleary, 1977: 432); in the example presented here, clients who are felt likely to fail a priori by program officials could be assigned to the "quick failure given ultimate failure" program to the benefit of the other program participants (Maltz and McCleary, 1977).

The use of time until failure measures has process implications as well. In the example concerning the evaluation of pre-trial intervention projects presented earlier, it was found through application of the Maltz-McCleary model that the median time until failure for ultimate rearrests was equal to 462 days; thus the one year data collection effort terminated before 50% of all ultimate rearrests had occurred! Again, had this calculation been performed, it could have been seen as a signal to prolong the data collection phase of the evaluation (Kaplan, 1978b:24).

To compare the timing of rearrests resulting from two programs with different client populations, the measure T_k (or $E(T_k)$ for Bayesians) is useful. One can examine the fractile times until failure to perform a relative comparison. Suppose we are interested in the time it takes until a fraction q of the population of ultimate failures has failed. For both of the programs being compared, compute

$$T_{k_i}|q \equiv \bar{n}^{-1}(N_i \cdot r_i \cdot q) \qquad i = 1, 2, \quad 0 < q < 1 \qquad (25)$$

for various values of q between 0 and 1. The values of $T_{k_i}|q$ (or $E(T_{k_i}|q)$) are then comparable for similar values of q (note that for $q = 1/2$, $T_{k_i}|q = t_{F(.5)}$, the median time until failure).

The final evaluation measure we will consider for the comparison of two programs is the safety time $t^*(\varepsilon)$. The values of t^* (or $E(t^*)$) as a function of ε may be plotted for each program on the same graph. It is then possible to check for dominance. Consider Figure 11 where $t^*(\varepsilon)$ has been plotted for two hypothetical programs. Here it is clear that Program A dominates Program B, since for any value of ε, $t^*_A < t^*_B$. To assert with confidence $(1-\varepsilon)$ that an individual will not be arrested given that (s)he hasn't failed by t^* will always require a longer time from release for individuals in Program B than for individuals in Program A. Of course, it is possible for partial dominance to occur; A could dominate B for low values of ε, while B could dominate A for high values of ε. For evaluation purposes, dominance over low values of ε characterizes a successful program.

It should be noted that in our discussion of the methods of this section, no specific form was assumed for $\phi(t|F)$. In fact, different hazard functions could be engaged for different programs, and the comparative procedures discussed here could still be invoked. Also, as mentioned by Maltz and McCleary (1977: 432), it is not necessary for programs to exist for the same length of time in order to use these techniques.

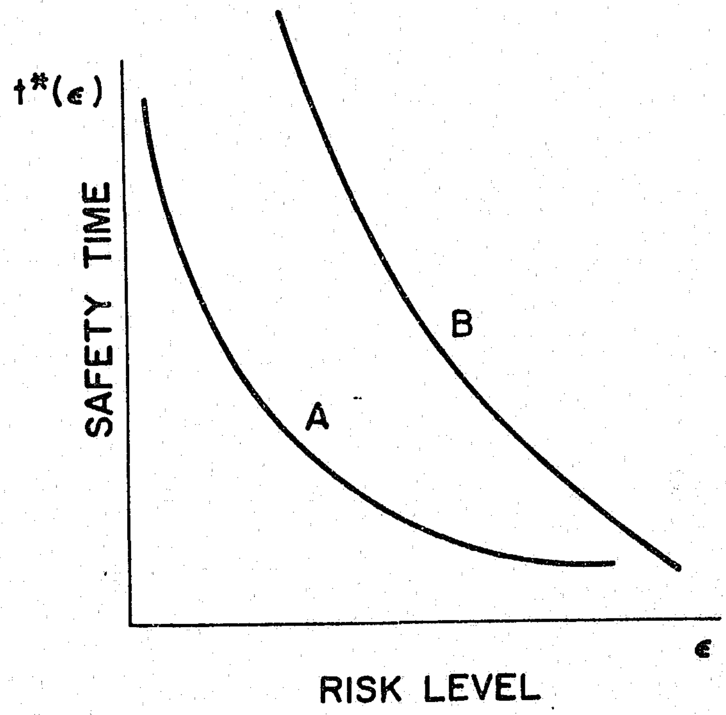


Figure 11

Program A Dominates Program B

VI. SUMMARY AND EXTENSIONS

This paper has analyzed the common structure shared by treatment-release corrections programs within the framework of reliability theory. Having presented a general model of the rearrest process, we examined the performance of this model under alternative assumptions, and illustrated appropriate techniques for estimating model parameters. We then discussed classical and Bayesian model-based evaluation procedures for use in both process and outcome situations.

While the substantive focus of this paper has been on models for rearrest patterns, it should be noted that the mathematics involved are appropriate for generic failure problems. Thus, if one was interested in performance measures based on alternative failure patterns over time, the reliability models of this paper could prove useful. For example, suppose one wished to judge a program participant as a failure only if that client was reconvicted. The time until failure would then correspond to the time until reconviction. Thus, the models presented here are responsive to the notion that different types of programs may require different definitions of client failure for evaluation purposes.

FOOTNOTES

[1]Reliability theory addresses problems associated with the failure of systems (or system components) over time.

[2]The pdf is a pseudo-pdf since $\int_0^\infty f_{t_F}(t)dt = r$, and r is in general not equal to one.

[3]While r is often considered to be a fundamental performance measure, the model presented by Blumstein and Larson (1971) suggests that 1/1-r is a more readily interpretable performance measure. The expression 1/1-r represents the number of future crimes committed per individual after release from treatment in the Blumstein/Larson model; this expression is very sensitive to changes in the value of r when r is close to one.

[4]It is assumed that the reader is familiar with the procedures of hypothesis testing; a good discussion of hypothesis testing is found in Chapters 10 through 12 of Freund (1971).

REFERENCES

1. Barnett, A. and Stabile, D., (1979). Estimating the "Virtual" Incapacitation Time for those Released from Corrections Programs. (Forthcoming).

2. Bloom, H.S., (1978). Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism. Discussion Paper D78-9, Department of City and Regional Planning, Harvard University.

3. Blumstein, A. and Larson, R.C., (1971). Problems in Modeling and Measuring Recidivism. Journal of Research on Crime and Delinquency 8: 124-132.

4. Campbell, D.T. and Stanley, J.C., (1966). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally.

5. Freund, J.E., (1971). Mathematical Statistics. Englewood Cliffs: Prentice-Hall, Inc.

6. Hillier, F.S. and Lieberman, G.J., (1974). Operations Research. San Francisco: Holden-Day, Inc.

7. Kaplan, E.H., (1978a). Evaluating the Effectiveness of One- versus Two-Officer Patrol Units. To appear in the Journal of Criminal Justice.

8. Kaplan, E.H., (1978b). Nine Pre-Trial Intervention Programs: A Re-Analysis. Unpublished paper, Department of Urban Studies and Planning, M.I.T.

9. Larson, R.C., (1975). What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment. Journal of Criminal Justice, 3: 267-297.

10. Maltz, M. and McCleary, R. (1977). The Mathematics of Behavioral Change: Recidivism and Construct Validity. Evaluation Quarterly, 1: 421-438.

11. Mullen, J., (Project Director) (1974). Pre-Trial Intervention: A Program Evaluation of Nine Manpower-based Pre-Trial Intervention Projects. Cambridge, Massachusetts: Abt Associates, Inc.

12. Purcell, E.J., (1972). Calculus with Analytic Geometry. Englewood Cliffs: Prentice-Hall, Inc.

13. Tsokos, C.P., (1972). Probability Distributions: An Introduction to Probability Theory with Applications. Belmont, California: Duxbury Press.

Appendix: Assymptotic Equivalence of the Bloom and Maltz-McCleary Models

As mentioned in Section II, Bloom (1978) has proposed a model involving an exponential hazard function. We also stated that Bloom's model was operationally equivalent to the Maltz-McCleary model. In this appendix, we will explain why this is so.

Bloom did not rely upon the notion of a conditional hazard function when formulating his model, as he rejected the assumption that some fraction r of the population released could be conceived *a priori* as consisting of ultimate failures (Bloom, 1978: 4). Rather, Bloom defined an *unconditional* hazard function for his model of the form (Bloom, 1978: 6)

$$\phi(t) = be^{-ct}, \; b, \; c > 0 \qquad \text{(A1)}$$

In Bloom's model, $\phi(t)$ represents the likelihood that an individual will fail in the next time instant given release at $t_R = 0$. Note that Bloom does not explicitly restrict the application of his hazard function to ultimate failures.

To obtain an expression for $p_F(t)$ given the hazard function of equation (A1), the well-known reliability result of equation (1) may again be invoked yielding (Bloom, 1978: 6)

$$p_F(t) = 1 - e^{-\frac{b}{c}} e^{\frac{b}{c} e^{-ct}} \qquad b, \; c > 0, \; t > 0 \qquad \text{(A2)}$$

This model has some interesting properties; foremost among these is the fact that by setting $t = \infty$ in equation (A2), one arrives at the expression (Bloom, 1978: 7)

$$p_F(\infty) = 1 - e^{-\frac{b}{c}} \quad b, c > 0 \qquad \text{(A3)}$$

Thus, Bloom's model implies that some fraction $1 - e^{-\frac{b}{c}}$ of the population released will ultimately fail. Bloom does interpret equation (A3) as an expression for the "ultimate probability of failure" (Bloom, 1978: 7), yet he rejects the notion that one may assume the <u>a priori</u> existence of an ultimate rearrest probability r. This distinction is at best artificial, as r and $1 - e^{-\frac{b}{c}}$ are both constants.

Recall that for the Maltz-McCleary model, $\phi(t|F) = a$, a positive constant. To see how the Bloom model asymptotically approaches the Maltz-McCleary model, we will formulate the <u>conditional</u> hazard function $\phi(t|F)$ for Bloom's model, and show that as t approaches infinity, $\phi(t|F)$ approaches a positive constant. If we define $p_F(t|F)$ as the probability of failure by time t given release at $t_R = 0$ and <u>ultimate failure</u>, then for Bloom's model,

$$p_F(t|F) = \frac{1 - e^{-\frac{b}{c}} e^{\frac{b}{c} e^{-ct}}}{1 - e^{-\frac{b}{c}}} \quad b, c > 0, \ t > 0 \qquad \text{(A4)}$$

Differentiating (A4) with respect to time from release to obtain $f_{t_F}(t|F)$, the conditional pdf for time until failure given ultimate failure, yields

$$f_{t_F}(t|F) = \frac{be^{-ct} e^{-\frac{b}{c}} e^{\frac{b}{c} e^{-ct}}}{1 - e^{-\frac{b}{c}}} \quad b, c > 0, \ t \geq 0 \qquad \text{(A5)}$$

Now, the probability that an ultimate failure will be rearrested in the next time instant conditioned on the event that (s)he has not failed by time t is given by

$$\phi(t|F)dt = \frac{\dfrac{be^{-ct}\, e^{-\frac{b}{c}}\, e^{\frac{b}{c}e^{-ct}}\, dt}{1 - e^{-\frac{b}{c}}}}{1 - \dfrac{1 - e^{-\frac{b}{c}}\, e^{\frac{b}{c}e^{-ct}}}{1 - e^{-\frac{b}{c}}}},$$

which algebraically reduces to

$$\phi(t|F) = \frac{be^{-ct}\, e^{\frac{b}{c}e^{-ct}}}{e^{\frac{b}{c}e^{-ct}} - 1} \qquad b,\ c > 0,\ t > 0 \tag{A6}$$

If we examine the limit of $\phi(t|F)$ as t approaches infinity, we realize that we cannot evaluate this limit directly since both the numerator and denominator of (A6) approach 0 as t approaches infinity. Applying L'Hôpital's rule (Purcell, 1972: 562), we obtain

$$\lim_{t \to \infty} \phi(t|F) = \lim_{t \to \infty} \frac{\frac{d}{dt}[be^{-ct}\, e^{\frac{b}{c}e^{-ct}}]}{\frac{d}{dt}[e^{\frac{b}{c}e^{-ct}} - 1]} \tag{A7}$$

$$= \lim\, [c + be^{-ct}]$$

$$= c \qquad\qquad c > 0.$$

Thus, the conditional hazard function for Bloom's model does approach a positive constant over time.

To illustrate the operational similarities between Bloom's model and the Maltz-McCleary model, we will return to our example of Section III. First, we impose the same restrictions on Bloom's model as those that were imposed on the Maltz-McCleary model:

$$\text{(i)} \quad p_F(6) = .2$$

$$\text{(ii)} \quad p_F(\infty) = .50 \ .$$

To satisfy restriction (ii), we use (A3) to obtain

$$b = -c \ln (.50). \tag{A8}$$

Substitution of (A8) into Bloom's expression for $p_F(t)$ given in (A2) combines with restriction (i) to produce the result c = .089. Placing this value for c in equation (A8) yields b = .062. These values of b and c may be used with equation (A2) to produce the results shown in Table AI. It is evident from Table A1 that these models behave in equivalent fashions.

Hence, it is not surprising that Bloom found the performance of his model and the Maltz-McCleary model to be operationally equivalent, despite their mathematical differences (Bloom, 1978: 16). When applied to the same data set, these two models will produce comparable results.

TABLE AI

FAILURE PROBABILITIES FOR THE MALTZ-MCCLEARY AND BLOOM MODELS: THE EXAMPLE OF SECTION III

Time from Release (months)	Maltz-McCleary ($r = .5$, $a = .12$)	Bloom ($b = .062$, $c = .089$)
0	0.00	0.00
6	0.25	0.25
12	0.38	0.36
18	0.44	0.42
24	0.47	0.46
30	0.49	0.48
36	0.49	0.49
42	0.50	0.49

END