harassing ex-offenders to encourage them to leave the jurisdiction. Or police may use computerized offender files when looking for suspects, and thus make more arrests of ex-offenders. A great deal of pressure is placed on the police to clear crimes by arrest, so they may be inclined to make arrests without sufficient cause. One is reminded of that memorable line uttered by Claude Raines, the impeccable chief of police in the movie Casablanca: "Round up the usual suspects."

Thus, using raw arrest data to indicate recidivism will produce Type I errors, i.e., including those who should be excluded. This will be the case insofar as the police arrest individuals who have not committed offenses. But Type II errors, errors of omission, also occur. In a great many cases people known to be guilty of a crime are not convicted, or ever arrested and for reasons totally unconnected with the quality of the evidence or the strength of the case:

- An offender may be put in a diversion program in lieu of prosecution.
- He may be granted immunity from prosecution in return for his testimony.
- His case may be continued so many times that witnesses die, move away, or just get discouraged from showing up for trial.
- He may offer to make restitution to the victim in return for the victim's agreeing to drop charges or withhold testimony.

- The charge on which the individual is convicted may bear little resemblance to the offense he committed due to plea bargaining.[3] An additional effect of plea bargaining is that the police may be inclined to "overcharge", that is, to charge an alleged offender with every possible offense in the hope that the bargain struck by the prosecutor will be close to the reality of the offense.[4]

Thus, the basic problem we are confronted with, even when we have complete information about offenders and their "transactions" with the criminal justice system, is the difference between de facto and de jure recidivism. Legal definitions of offenses, arrests, and charges are the basis of the only data we can obtain directly from the criminal justice system, and must be used by evaluators in assessing behavioral characteristics. Were the legal definitions reflective of what actually occurred, the problem would be minimized. However, this is not the case -- the data are distorted because sentences are based on these definitions of criminal conduct: A buglery may be charged as burglary, criminal trespass, malicious mischief, or larceny, which is not very helpful when one is trying to reconstruct what happened.

We see, then, the problems associated with operationally defining recidivism by using criminal history records, even when the records are complete. On the one hand we have errors of commission if we call an arrestee a recidivist when he has been arrested without having

committed an offense. On the other hand we have errors of omission if arrestees who are factually guilty are labeled non-recidivists, because they have not been convicted for the variety of reasons discussed above. Based on the empirical data relating to these errors, Blumstein and Cohen (1980) have concluded that "the errors of commission associated with truly false arrests are believed to be far less serious than the errors of omission that would occur if the more stringent standard of conviction were required." The weight of the empirical evidence also inclines us toward this point of view. However, it may also be advisable to use some sort of check on the quality of the arrest. We discuss this issue later in this chapter, in describing our suggestions for definitions of recidivism events.

Problems with Incomplete Criminal Justice Data

In the last section we assumed that complete criminal justice information is available to us. However, ccmplete information about an individual's transactions with the criminal justice system may be lacking. Many different agencies report on these transactions, but in most states there is no single central repository for the data. This is in part attributable to the many jurisdictions and governmental levels involved in the criminal justice process in a state.

Enforcement data. A state's enforcement agencies are found at all governmental levels: municipal (local police), county (sheriff's

offices), and state (highway patrols, bureaus of investigation). Centralized reporting of crime and arrest data was initiated over forty years ago and has been improved and expanded over the years (Maltz, 1977).

All enforcement agencies report felony arrest data to the FBI's National Crime Information Center (NCIC). In the past these reports went directly to the FBI; but recently many states have established Statistical Analysis Centers (SACs), funded by LEAA, to compile these data and forward them to the FBI. These states are able to provide researchers with virtually complete felony arrest data.

Prosecutorial and court data. With few exceptions, criminal courts and prosecutor's offices are county agencies. There is no single repository for prosecution and trial data akin to the NCIC Program for felony arrests. SACs are charged with collecting all criminal justice data, including prosecutorial and trial data, but this aspect of their work has not proceeded as rapidly as has the collection of data from enforcement agencies. However, many of the larger prosecutor's offices have been funded by LEAA to install the Prosecutor's Management Information System (PROMIS), or related computer-based information storage and retrieval systems. Data from these offices should be more complete than can be anticipated from other jurisdictions.

Correctional data. Correctional data are generated at both the county (jail) and state (prison, halfway house) level. But few states presently include jail data in their statewide correctional statistics. Here, too, developments are promising. The LEAA-funded Offender-Based State Correctional Information System (OBSCIS) will compile state correctional data at all levels. At the present time, however, few states can routinely track individuals who have been given jail sentences with any assurance of completeness.

So we see another reason for using the arrest as the indicator of choice for recidivism. Prosecutorial, court, and correctional data are not yet as complete or reliable as arrest data supplied by enforcement agencies.

State-specific data. Even if all of the information from within the state were available for analysis it would still be incomplete for offenders who are rearrested, prosecuted and/or convicted in another state. If only data from the state conducting the study are used -- this is always so when "return to prison" is the recidivism indicator -- recidivism rates may be dependent on geography. For example, state-specific recidivism rates for states like Rhode Island and Nevada may be lower than expected due only to the proximity of their major population centers to bordering states. States have different regulations limiting the dissemination of such data for interstate (and intrastate) use (Office of Technology Assessment, 1978), so one cannot count on their availability.

Data Problems in Controlling for Arrest Quality

Recidivism need not be determined by arrest alone. One can attempt to reduce type I errors (improper arrests) by looking at subsequent criminal justice transactions based on the arrest. If recidivism is defined as arrest only if followed by affirmative prosecutorial action, records must be examined to see if an indictment, information or other prosecutorial action has taken place. To examine these records the analyst must go through all files in a prosecutor's office. Except for the few offices that have been automated, this usually means that every manila folder in the office must be examined to determine the name of the defendant, the nature of the original charge at arrest, whether the charges have been dropped or pursued further, and the present disposition of the case. This information is usually handwritten on the outside of the folder by staff attorneys whose penmanship rivals that of the medical profession for illegibility and encryption. In other words, this task is not quite so straightforward as it seems, especially since many of the folders are likely to be in brief cases, in piles on (and under) desks, in attorneys' homes, in their cars, or in other locations convenient for the attorneys but not the researcher. Although this particular check on the quality of an arrest would be useful, one cannot be sure of getting complete and accurate data.

Recidivism may also be defined as arrest only if followed by conviction. Determining if a conviction ensued is less difficult a task than determining the nature of the prosecutor's response. The FBI's National Crime Information Center compiles data on arrests and subsequent dispositions. However, the NCIC program is largely voluntary, so complete disposition information cannot be guaranteed. But the quality of dispositional data is improving rapidly, especially in those states which have Statistical Analysis Centers or other central repositories of criminal justice data. For the time being, though, the researcher studying recidivism -- regardless of the definition used -- should not expect to have complete data available.

Absconsion

Another disposition related to the parole process must be considered in the definition of recidivism. If a person absconds, that is, stops reporting to his parole officer and cannot be found, should he be considered a recidivist? In some states, strangely enough, an absconder is treated as a success -- by default. If the absconder has no action taken against him because he cannot be found, he is carried on the books as a success because he has had no parole violation, arrest or other adverse action taken against him. Even if this oversight is corrected, the question of how to treat absconsions is still not answered. No doubt many parolees abscond because of fear of

arrest, but absconsions should not automatically be treated as failures. How they are treated may depend upon the way failure is defined.

Recidivism Defined in Practice

Most states only use data from their own state for evaluative purposes. This reliance is due more to the uncertainty and difficulty of obtaining data from other states than to any theories about offender mobility. For special evaluations FBI data may be used in follow-ups, so that all arrests are included (Kitchener, Schmidt, and Glaser, 1977; Hoffman and Stone-Meierhoefer, 1977), but this is not the rule. And even when FBI data are used, it is often very difficult to determine if a conviction resulted from an arrest: Kitchener et al (1977) found this to be so in their study of releasees from federal prisons, even though they had the FBI's NCIC data available to them; and Hoffman and Beck (1974) found that disposition data were often missing.

But arrests, whether or not followed by convictions, are not the only indicators of recidivism that are used in evaluations. For example, Waldo and Chiricos (1977) used eighteen different measures of recidivism in a study of a work release program. In order to determine the way recidivism has been defined in practice, we reviewed some ninety studies which used recidivism as an outcome measure. The

studies were in the open literature and/or cited by Lipton et al (1975). We separated the definitions into nine general categories. These categories and some of their qualifying conditions are:

- o Arrest -- number of arrests; recorded police contact; court appearance; time elapsed before the first arrest; did conviction result?
- o Reconviction -- jail or prison sentence; felony or less; sentence;
- o Incarceration -- type of facility; seriousness of offense;
- o Parole violation -- nature of the violation; seriousness of the infraction; police initiated;
- o Parole revocation -- new offense; seriousness of the offense; average number of good days on parole;
- o Offense -- seriousness; number; new offense;
- o Absconding -- was an absconder warrant issued;
- o Probation -- proportion redetained; length of time detained; number of violations; violation warrant; incarceration.

Each of these definitions has been used in research. Examples include:

Arrest. Arrest for a new offense (Inciardi, 1971; Fishman, 1978; rearrest (Cox, 1977); court appearance within one year after release (Coates, Miller & Ohlin, 1977); criminal arrest (Hoffman & Stone-Meierhoefer, 1977); recorded police contact (Wolfgang, Figlio & Sellin, 1972; Murray, Thomson & Israel, 1978); number of arrests (Waldo & Chiricos, 1977).

Reconviction. Reconviction of a felony (Hopkins, 1976); reconviction or recall for unsatisfactory conduct (Hood, 1966); any new conviction resulting in a sentence of sixty days or more (Gottfredson & Ballard, 1965; Bennett & Ziegler, 1975; Beck & Hoffman, 1976); any new conviction for a felony or felony-like offense (Kitchener, Schmidt & Glaser, 1977); conviction of a further offense (Wilkins, 1958).

Incarceration. Jail sentence of more than three days and return to prison for either a new offense or a technical violation (Burkhart & Sathmary, 1964); return to prison (Jacobson & McGee, 1965; Waldo & Chiricos, 1977); return to prison as a parole violator or outstanding absconder warrant (Gottfredson & Ballard, 1965); number of commitments to an institution with or without adjudication (Boston University, 1966); return to prison for an administrative violation (Beck & Hoffman, 1976); recommitted to prison with conviction(s) for major offense, same jurisdiction or any other jurisdiction (Mosely & Gerould, 1975); return to prison for at least thirty days within one year of release (LeClair, 1977); return to prison as a parole violator (Kitchener, Schmidt & Glaser, 1977).

Parole Violation. Alcoholic criminality: arrests and fines for drunkenness, disorderly conduct or nonalcoholic criminality: more serious offenses (Hansen & Teilman, 1954); issuance of a parole viola-

tion warrant whether subject was reinstitutionalized or not (Garrity, 1956); violating rules of parole (Kusuda & Babst, 1964); issuance of a parole violation warrant by District Court or Board of Parole for technical violations or new felony offenses (Lohman, 1967); parole violation (Kantrowitz, 1977); parole violation warrant (Hoffman & Stone-Meierhoefer, 1977).

Parole Suspension. Parole suspension with or without a new offense (Narloch, Adams & Jenkins, 1959); suspension (California, 1956); number of parole suspensions (Werner & Palmer, 1976).

Parole Revocation. Revocation of parole (Johnson, 1962); parole revocation or bad discharge (Guttman, 1963).

Offense. New offense, or violation rate: defined as violation of the rules of supervision within two years (Babst & Mannering, 1965); mean number of offenses during the twelve-month follow-up period, seriousness of the offense (McEachern & Taylor, 1967); Waldo & Chiricos, 1977).

Absconding. Absconding (Kusuda & Babst, 1964; Inciardi, 1971; Mosely & Gerould, 1975); absconder warrant (Gottfredson & Ballard, 1965; Beck & Hoffman, 1976).

Probation. Observed reconviction rate compared with expected reconviction rate (Great Britain, 1964); drunk arrest rate for municipal court (Ditman & Crawford, 1965); unfavorable dismissal from probation (Feistman, 1966); number of months before successful completion of probation (Kowaguchi & Siff, 1967).

As can be seen, in some studies more than one definition (e.g., "either parole revocation or arrest") was used. Table 6-1 was compiled from about ninety studies listed in the Bibliography and denoted by an asterisk. The figures should be taken as indicative of the popularity of different measures rather than as an authoritative exact count.

There can be major differences in the conclusions about correctional effectiveness, depending upon which recidivism measure is used.[5] Hoffman and Stone-Meierhoefer (1980) have shown how varying the way recidivism is operationalized can produce different interpretations of which program or cohort is better. Obviously, what is needed is a way to systematize the operational definitions of recidivism.

Proposed Recidivism Definitions

We do not propose to prescribe one single operational definition of recidivism. Rather, we propose to describe some of the more useful definitions in some logical way. Our goal is similar to the goal of economists in developing a set of definitions of "money." Economists talk not of money in general, but of M1 -- currency + demand deposits -- or M2 -- M1 + time deposits other than large certificates of deposit. These have been found to be useful indicators in charting the economy's course. We suggest below a similar language for recidivism.

TABLE 6-1

RECIDIVISM DEFINITIONS USED IN RECENT STUDIES

Definitions	Frequency
Offense data	
Recorded police contact	2
New offense	16
Severity of offense	12
Arrest	20
Parole/probation infractions	
Parole suspension	8
Parole revocation	8
Technical violation	26
Absconding	10
Probation violation	3
Court appearance	3
Reconviction	22
Sentencing	8
Return to prison	39

Source: Compiled from over ninety studies listed in Bibliography

Recidivism is normally measured in terms of the time interval between two events: time of release and time of recidivism. (Of course, those who do not recidivate do not experience the second event.) For example, many studies of correctional programs report the "one-year recidivism rate," which is the fraction of program participants who experience a recidivating event within one year of release.

The release event can be one of a number of events: release from incarceration, release from parole supervision, release from a work-study program or halfway house. The choice of event can depend on the nature of the program being evaluated. For instance, to evaluate a prison vocational program one would consider release from prison to be the release event whether or not the participants were on parole.[6] To evaluate a parole program the release event would be release from parole supervision.

Occasionally it may be necessary to look beyond recorded dates to determine when the release event occurred. For example, when a prison releasee enters a halfway house he may find himself essentially incarcerated for the first few months, until he gets his bearings in his new surroundings. (He may be permitted out to work, but never without supervision.) In this case the actual date of release may not be the date he leaves prison, but the date he is released on his own to the outside world.[7]

The recidivism event depends upon the nature of the release event. If the release event was release to parole or probation supervision, the recidivism event may be a technical violation (of the conditions of parole or probation or of the rules of the halfway house); it may be revocation of parole or probation in lieu of trial for a new offense; it may be arrest for a new offense (whether or not followed by an additional disposition); or it may be return to prison.

Not all of the possible release/recidivism pairs will be defined. Our concern has been to operationalize recidivism in ways that are reasonable from theoretical and data acquisition standpoints. The definitions are based on the events that terminate the time interval, because the date of release (when the recidivism clock starts) is both less ambiguous and more dependent on the program under evaluation. They are listed in order of most restrictive to least restrictive.

R_{AC} -- arrest and conviction. The time interval runs from date of release to date of arrest, but it is counted as a recidivism event only if the arrest results in conviction. An absconder is treated as an as-yet non-failure up until the date of absconsion, except if he absconded to avoid arrest.

$R_{AC/s}$: Same as R_{AC}, but using data from only one state.

R_{AP} -- arrest and prosecution. The time interval runs from date of release to date of arrest, but it is counted as a recidivism event only if some prosecutorial action is taken against the arrestee: charges filed, grand jury presentation, indictment, etc. An absconder

is treated as in the previous case.

$R_{AP/s}$: Same as R_{AP}, but using data from only one state.

R_{A} -- arrest only. The time interval runs from date of release to date of arrest, regardless of whether conviction ensues. An absconder is treated as in the previous case.

$R_{A/s}$: Same as R_{A}, but using data from only one state.

R_{VC} -- violation and return to custody. The time interval runs from date of release to date of violation of the terms of release (i.e., parole or probation), but it is counted as a recidivism event only if the violator is returned to custody. Violations can include return to custody in lieu of arrest for another offense.[8] Absconders are treated as having failed on the date of absconsion.

R_{V} -- violation. The time interval runs from date of release to date of violation of the terms of release, whether or not the individual was returned to custody. For this category the types of violations which are to be defined as constituting recidivism should be specified, so absconders may be treated as failures or as as-yet non-failures.

Another measure of recidivism that has been used is "return to prison". The time interval usually runs from date of release to date of return to prison, not to the date of the arrest that led to the prison sentence. This measure may be useful to prison officials, for planning budgets and expected occupancy rates. However, it is <u>not</u>

useful as an indicator of offender behavior because it includes criminal justice processing time. The time interval is the sum of the following time intervals: release to arrest; arrest to hearing; hearing to trial; trial to sentencing; and sentencing to recommitment. Only the first time interval relates to offender behavior; the others reflect the behavior of the criminal justice system.

To summarize, the recidivism definition of choice appears to be R_A, arrest recidivism, for the present. [For studying parolee or probationer cohorts R_V would appear to be most appropriate, since technical violations frequently mask new offenses.] R_{AP} would be preferred were there data to support it, but this is not presently the case; therefore, the choice is between R_{AC} and R_A. And despite the fact that R_{AC} includes arrests for which no offense occurred, their number is (probably) considerably smaller than the number of felony arrests of guilty persons for which no felony conviction resulted. Moreover, arrest data are more accessible than disposition data. The choice is dictated, then, by data availability and completeness, not by theoretical considerations.

Other Sources of Noncomparability

We do not mean to imply that use of the same recidivism measurement by different programs or jurisdictions will automatically result in comparable findings. This is far from the case. Differences in laws, policies and procedures among jurisdictions will prevent direct comparisons. As an example, suppose one state makes little use of probation while another uses it heavily. Then a cohort of prison releasees in the first state will probably have a higher percentage of good risks than will the latter state, and therefore, a lower recidivism rate. Or suppose one jurisdiction has a computerized data collection system used by parole officers while in another jurisdiction parole officers turn in longhand reports on their parolee caseload. Then information about technical violations is likely to be considerably more variable in quality (and often richer) in the latter case than in the former. Not all of these differences can be accounted for; foreknowledge of their presence and their potential effect on recidivism measurement is necessary in developing evaluation designs and interpreting results.

NOTES

1. Robbery is usually categorized as a personal crime, not a property crime. The reason is that the victim has been placed in danger by the offender. But in studying recidivism we are concerned with the offender's behavior, not the victim's. Since robbery is more an instrumental crime -- to obtain money -- than an expressive crime -- to fulfill psychic needs -- it would appear that someone who engages in robbery would have a greater similarity to someone who engages in property crime than with someone who engages in murder, rape or assault. In fact, Sokal (1974: 1121) found that robbery was more strongly associated with auto theft than with any other index crime.

2. One might also consider using prosecutorial action as a quality control check on arrests. That is, if the prosecutor presses the charges, one can assume that the arrest is valid. However, obtaining data from this part of the criminal justice system is often more difficult than obtaining data from police, courts and correctional agencies. This problem will be discussed later in this chapter.

3. Determinate, or "flat time" sentencing (Fogel, 1979), which limits the sentencing discretion of judges, is becoming more and more widespread. One consequence may be an increase in plea bargaining by prosecutors, effectively moving sentencing discretion to this stage of the criminal justice process.

4. In some jurisdictions (including Cook County, Illinois) the prosecutor's office has a felony review unit. The unit is informed whenever a felony arrest is to be made and insures that there is sufficient evidence to warrant a felony arrest. This minimizes the problem of overcharging. (And by weeding out all of the hard-to-try cases, this procedure also has the benefit of improving a prosecutor's win-loss record.)

5. The Uniform Parole Reports Program presently collects parole data from all states without distinguishing among them based on their parole organization and structure. Conclusions drawn from such data may be misleading. See Chapter 9 for an analysis of UPR data.

6. Naturally, one would control inter alia for parole participation in the program evaluation.

7. It may be worth considering the extent of a person's freedom as a variable of interest. Living alone, a person has a great deal more free time (and a greater risk of recidivating) than when living in a supportive (and confining) environment such as a halfway house.

8. Hoffman and Stone-Meierhoefer (1977) use this criterion, but add "killed during the commission of a crime." Naturally, this event should be considered a recidivism event, but it may not always be recorded. In any event, it occurs so infrequently that its exclusion (should the data not be available) would not greatly affect the results.

CHAPTER 7

SETTING THE STAGE FOR DATA ANALYSIS: UNDERLYING ASSUMPTIONS AND MODELS

Consistency in defining recidivism permits one to compare different groups of releasees, and to determine the statistical significance of differences that may arise. If the groups represent different offender types, e.g., twenty-year-olds versus thirty-year olds, one can then make statements about the effect of age on recidivism. If the groups are being compared to test different treatments, then one assumes that the difference in recidivism is not due to the composition of the groups. But even when the same selection criteria are used for all groups, this assumption may not hold. Other assumptions relate to the characteristics of the recidivism process, i.e., the "model" of recidivism used. These assumptions are discussed in this chapter.

Comparing Groups

A consistent definition of recidivism is not by itself sufficient to insure comparability. As should be evident from the previous chapter, variations in state laws and policies on parole make state-to-state comparisons[1] unreliable, even if we were to assume that the population characteristics of all states were similar.

But populations are not similar, from state to state or from institution to institution within a state. Even when studying a single subpopulation (e.g., all persons released from Joliet, aged 25-35, who were convicted of armed robbery) a great deal of variation can exist between those selected for an experimental program and those who constitute the control group. One way to compensate for such variation is to randomly assign individuals who meet the program criteria to experimental and control groups. Although this does not insure that the groups will be similar,[2] the chances of obtaining similar groups are enhanced considerably.

Random assignment procedures are helpful in searching for causal links. They are frequently used, for example, to study the effect of drugs on a standardized strain of laboratory animals, with the animals assigned randomly to experimental and control groups. But when dealing with social programs the researcher must contend with great variability within the groups under study. This can be accomplished by increasing the size of the groups, so that within-group differences are "averaged out" and between-group differences can be discerned. The cost of the study, of course, increases as a consequence of this variability.

A more serious problem is the difficulty in getting random assignment procedures accepted in studying social programs. Program administrators are loath to permit (or let it be known that they permit)

the assignment of persons to Program A or Program B based on the flip of a coin;[3] judges reject the idea of sentencing people using a table of random numbers.[4] In many cases ethical considerations dictate that people volunteer for the experimental program. Moreover, a judge will sentence people to programs based on his/her perception of whether the program will benefit them.[5] So random assignment is difficult to attain in studying social programs.[6]

Another factor militating against true experiments in social programs is the inability to disguise the experimental treatment. In assessing the effectiveness of fluoride toothpaste one can use a double-blind experiment, so that neither the toothpaste user nor the dentist knows whether a given person is in the experimental or control group. An innovative vocational training program cannot be similarly disguised: the participants know whether they drew the program or the control, and this knowledge may affect their performance or the expectations of the program staff and researchers.

Other problems in implementing a random selection procedure are described in Cook and Campbell (1979: Chapter 8), as are situations conducive to randomized experiments in studying social programs. The alternative to an experimental design is to use quasi-experimental designs. These have been used to a great extent in correctional evaluations, but as Cook and Campbell point out (Chapter 3), ensuring comparability among groups is an even greater problem.

Modeling the Recidivism Process

The remaining assumptions discussed in this chapter concern the characteristics of the recidivism process. Two important structural features are included in our characterization: a) the measure of program effectiveness is binary (success or failure); and b) the time to failure is the primary outcome variable of interest. Although success and failure are of course actually multi-valued attributes, we assume that a binary variable is a good description for policy purposes.[7]

Test of Proportion

The most common method of measuring recidivism is to fix a point in time -- often one or two years -- and determine the fraction of participants in each program who have failed within that length of time. This is obviously not the total fraction of failures, since others will undoubtedly fail after this time: rarely is the time interval long enough to observe all eventual failures.[8] A statistical test is then used to determine whether there is a significant difference between the two proportions of failures.

The proper use of the difference of proportions test is when both groups are samples of larger populations. The question asked is, "Is the probability less than .05 that the two parent populations [from which the samples were drawn] have the same proportion of failures, or

is the difference in proportions is merely due to chance?" In other words, the difference of proportions test uses the properties of samples of populations to test the characteristics of their parent populations.

In a correctional program evaluation one does not ordinarily administer a program treatment to an entire population, and only then draw a sample; the sample is drawn first and then the treatment is administered, but only to the sample. One must therefore consider whether the program is expected to have the same effect when it is administered to the entire population rather than just the sample.

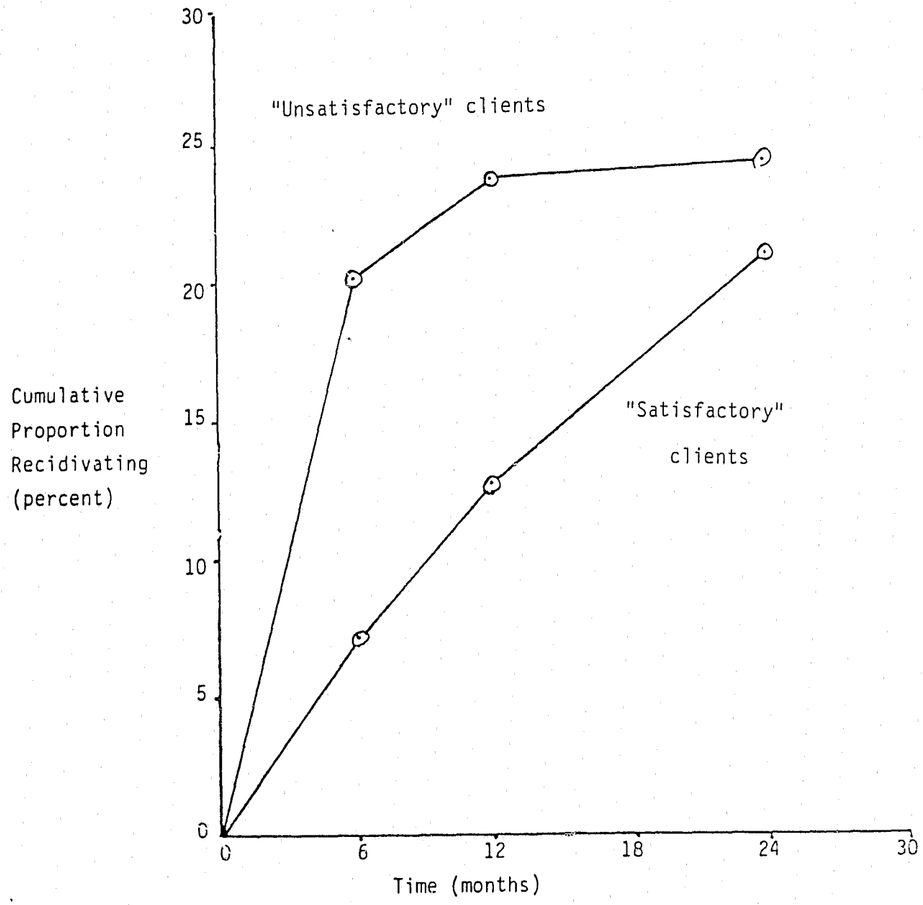
If the program can be "scaled up" from the sample to the population then the two situations are equivalent. This may be the case for, say, evaluations of new drugs -- one would expect them to have the same effect on the parent population as on a properly drawn random sample of the population. However, this is not generally true for social programs: as the word "social" implies, it is the interaction among the actors -- program participants and staff -- that often constitutes (or enhances) the treatment. And the experience of the past decade leads us to conclude that social treatments which are effective on the experimental, pilot level do not work as well when they are bureaucratized and/or extended to a larger population.[9]

But this means of measuring recidivism has an even greater problem. One must fix a point in time at which to calculate the

difference of proportions. This point in time is arbitrary, and the method itself ignores the time-varying nature of the failure process. Consider the proportions of failures for the two cohorts shown in Figure 7-1. The data are from Minnesota (1976: 196). As can be seen, the "satisfactory" clients had lower recidivism rates at six, twelve and 24 months (data were not collected at 18 months). If one were to project beyond two years, however, it is likely that the "unsatisfactory" clients would ultimately have a lower percentage of recidivists. Looking at only one of these three points in time does not tell the whole story.

In fact, there is no agreement on the most appropriate follow-up time for studies of offender post-release behavior. Most correctional studies use the "one-year recidivism rate", without any justification (other than, perhaps, the fact that it coincides with the funding cycle). The National Advisory Commission for Criminal Justice Standards and Goals, on the other hand, has recommended that recidivism should be based on a three-year follow-up period (NAC, 1973: 93). There should be a more rational means of deciding the length of follow-up time.

The difference of proportions test, then, is based on a model which is inappropriate to studying recidivism. This leads us to investigate alternative models.



Source: Minnesota, 1976: 196

Figure 7-1 Recidivism Rates of Two Groups of Clients Released from Minnesota Halfway Houses

Other Models of the Recidivism Process

The alternative models we consider treat the groups under study not as samples, but as the whole populations under consideration. Statistics are used to estimate the parameters of the model, and not the characteristics of a (nonexistent) parent population.

A number of such models have been suggested and applied in recent years: (Stollmack & Harris, 1974; Harris & Moitra, 1978; Witte & Schmidt, 1977; Maltz & McCleary, 1977). Their characteristics are discussed below.

The Exponential Distribution

Stollmack and Harris (1974) broke new ground in the analysis of correctional programs. Rather than looking at the proportion of failures after a given time interval (usually one year), they suggested looking at recidivism as a process in which the time to failure of an individual is considered to be a random variable. Such a failure rate model is well-developed in the study of reliability of machinery (Barlow & Proschan, 1975) and the study of biomedical problems (Gross & Clark, 1975). One then considers the probability that an individual will fail before a given time. When the failure rate is constant and does not depend on time, the probability of a failure before a given time is exponentially distributed, as shown in Figure 7-2.

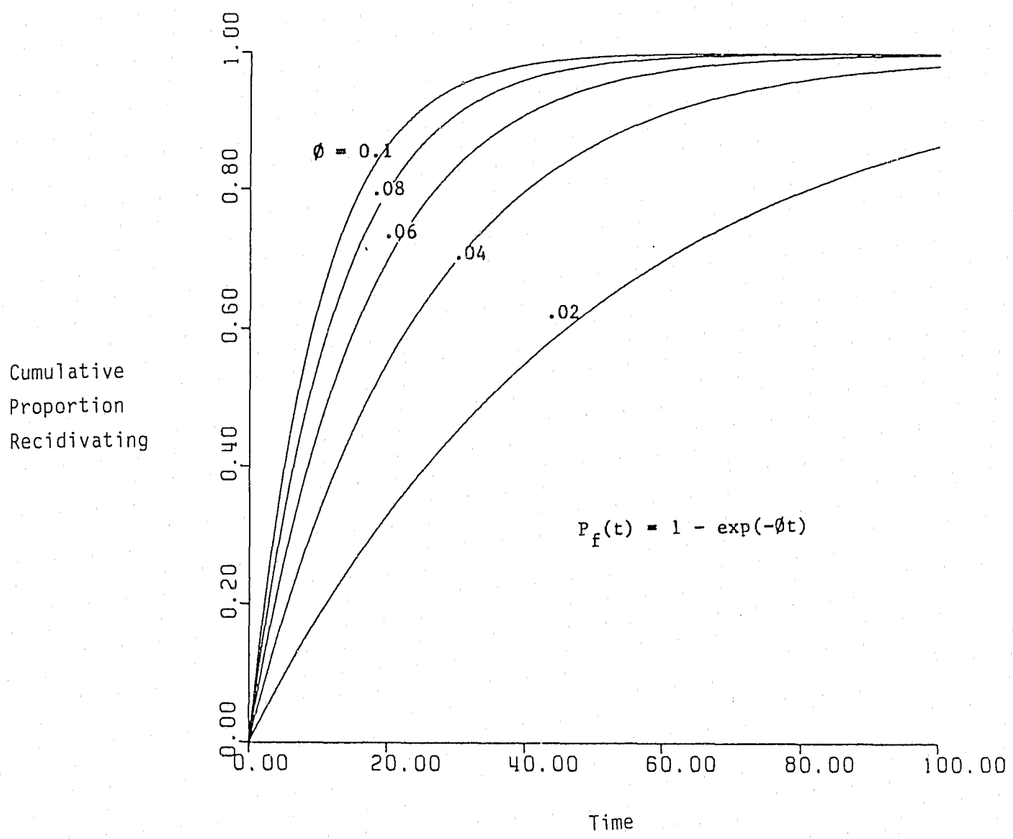


Figure 7-2 The Exponential Distribution

In their numerical example, Stollmack and Harris (1974) used such a model in their analysis of a halfway house program in Washington, D.C. They first used the conventional difference of proportions test, which failed to show any significant difference between experimentals and controls. Using the failure rate model, however, they found a significant difference between the failure rates of the two groups.

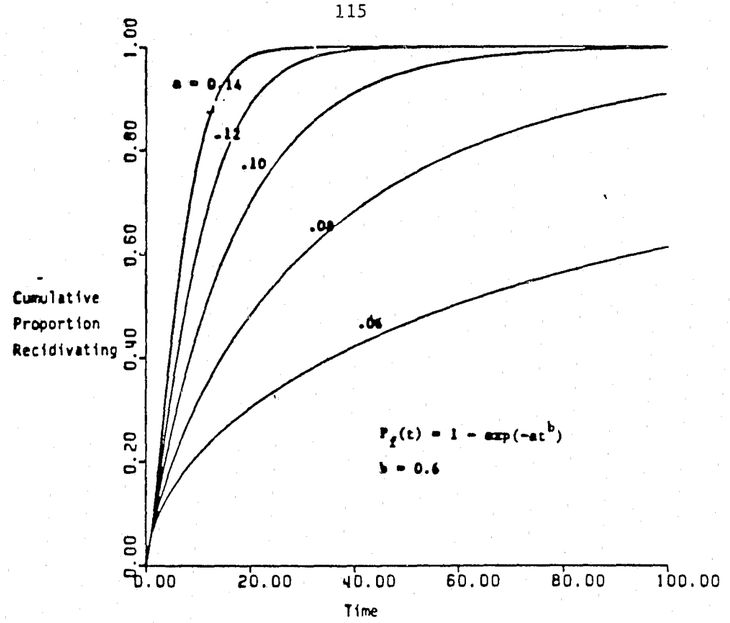
The Weibull Distribution

Harris and Moitra (1978) have used a non-constant failure rate model, leading to a Weibull distribution of failure times, to analyze recidivism data. The Weibull distribution requires two parameters (rather than the single parameter of the exponential model). Figure 7-3 shows how the two parameters affect the shape of the cumulative distribution of failure times.

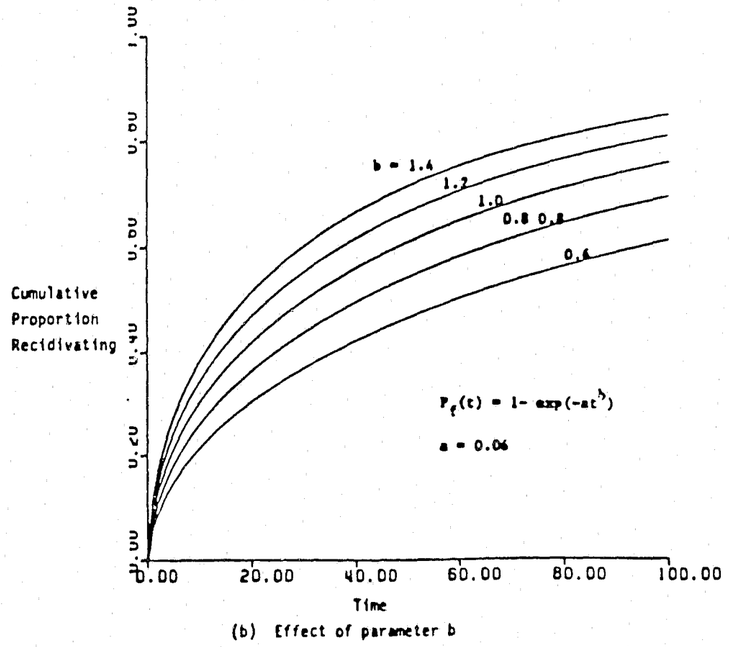
The Weibull distribution has the advantage that it can be used to model cases where failure rates, may increase or decrease over time. (Recall that the exponential model assumes a constant failure rate over time.)

The Lognormal Distribution

Witte and Schmidt (1977) analyzed recidivism data for a sample of individuals released from prison in North Carolina. They used the lognormal distribution for time to failure "because it fits the data



(a) Effect of parameter a



(b) Effect of parameter b

Figure 7-3 The Weibull Distribution

fairly well; better than the exponential or normal distribution, for example." Like the Weibull distribution, the lognormal distribution requires two parameters. The effect of the two parameters on the shape of the distribution is shown in Figure 7-4. The failure rate for the lognormal distribution first increases, then decreases over time.

The Split-Population Distribution

Another two-parameter distribution of failure times that can be used to analyze recidivism data is one we have termed the split-population distribution. The model is similar to the one based on the exponential distribution, with one important exception: it does not assume (as the exponential model does) that all program participants will eventually fail. The reason for this situation is understandable when one looks into the origin of failure rate analysis.

As previously mentioned, failure rate analysis has its roots in studies of the reliability and lifetimes of electromechanical components. The failure of such devices is assured: Eventually they will all wear out, so it is unnecessary to include the possibility of success in the model.

However, no correctional analysis should be based on a model which assumes a priori that failure is inevitable, that everyone will eventually fail -- whether or not it ultimately turns out to be true. As

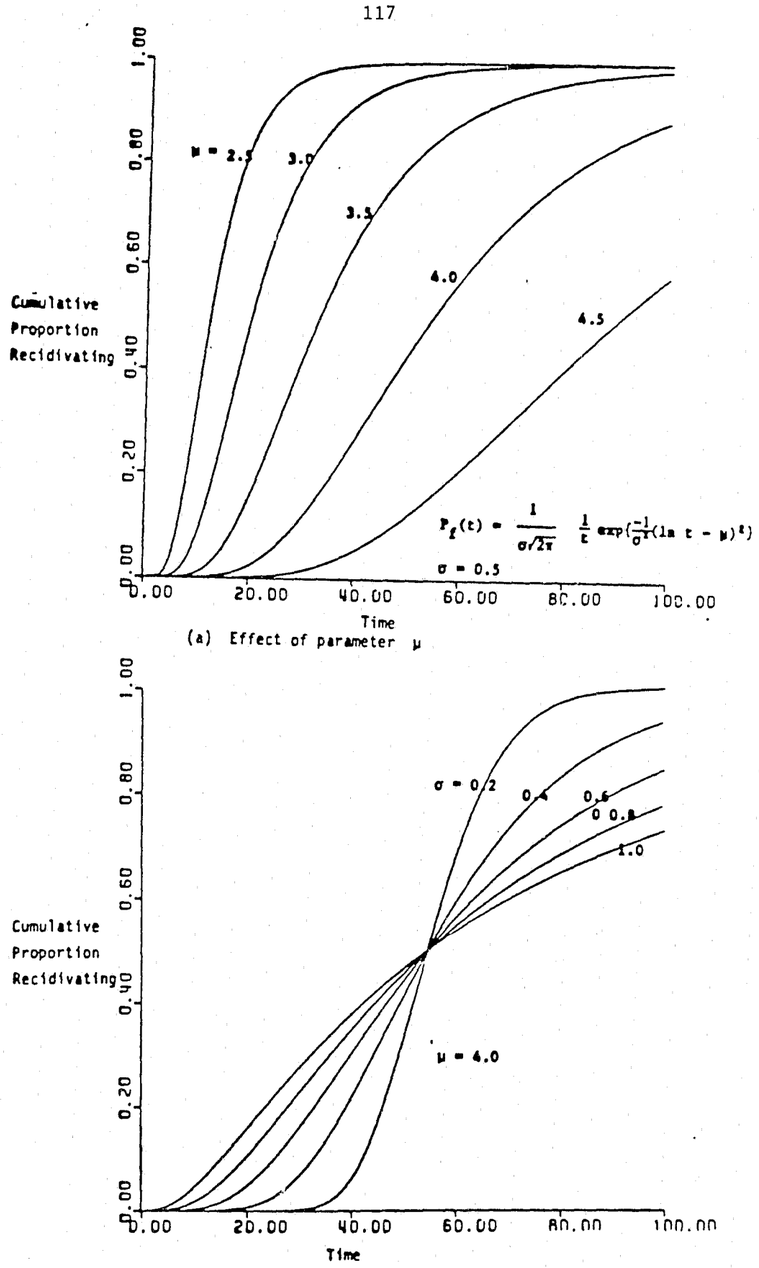


Figure 7-4 The Lognormal Distribution

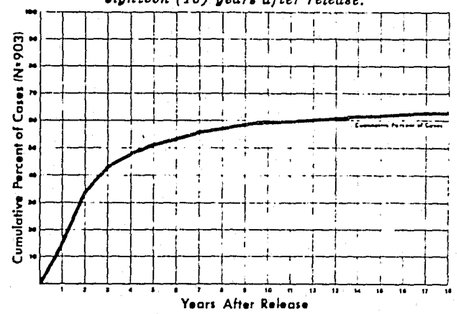
can be seen from Figure 7-5, which shows the results of a long-term follow-up of offenders released from federal prisons, failure is not inevitable. The exponential, Weibull and lognormal distributions (Figures 7-2 to 7-4) all approach 1.0 as time increases; this is tantamount to assuming that everyone eventually fails.

The split-population model assumes that the group under study consists of individuals each of whom has a non-zero probability of succeeding. Those who fail are assumed to do so at a constant failure rate, resulting in a cumulative probability distribution as in Figure 7-6.

Two parameters characterize this model: the probability that an individual member of the group will eventually fail, and the (constant) rate at which the eventual failures do fail. These two parameters can be estimated without too much difficulty (Maltz & McCleary, 1977; Lloyd & Joe, 1978).

This model is based on a frequently used analytic device in the social sciences. When a single-population model fails to fit a social phenomenon, analysts have employed the next simplest assumption -- that a population dichotomy results from the phenomenon. Partanen (1969) was the first to employ it in studying recidivism. The mover-stayer model used in demographic studies (Goodman, 1963) and the two-fluid model of town traffic (Herman & Prigogine, 1979) are examples in other contexts. Similar models have been used to study other

FIGURE 1.—*How many fail within eighteen (18) years after release.*



Source: Kitchener et al, 1977

Figure 7-5 An Empirical Basis for the Split Population Distribution

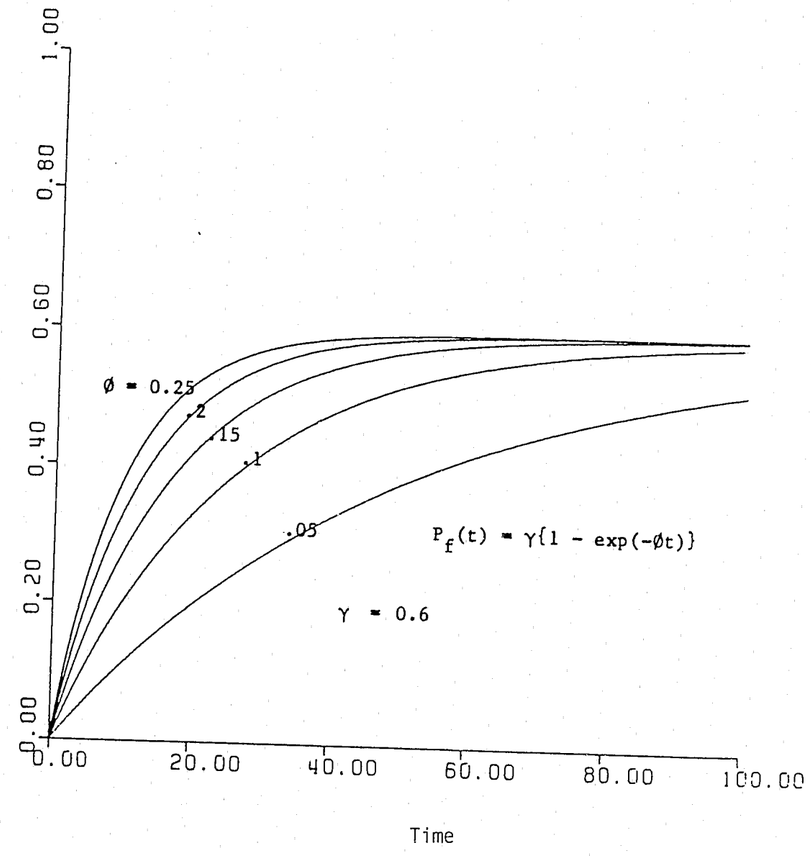


Figure 7-6 The Split Population Distribution

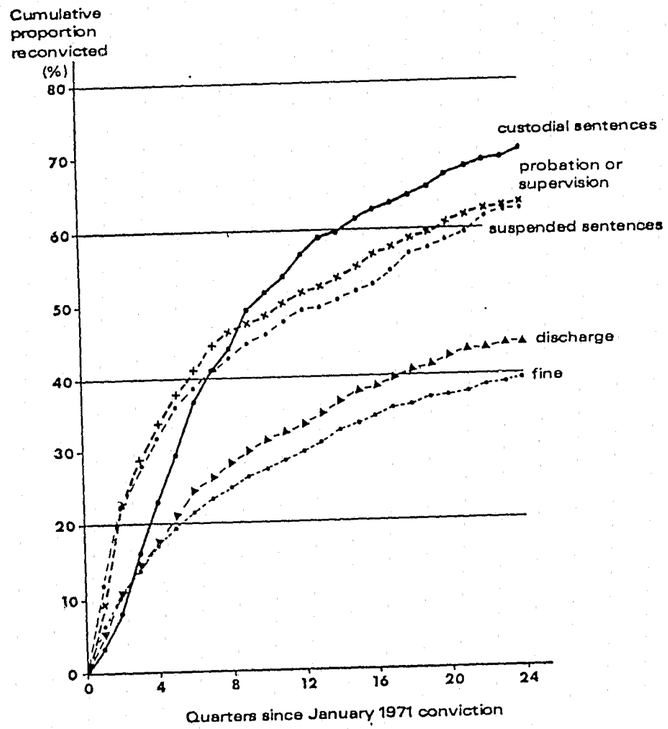
failure processes, including cancer mortality (Boag, 1963; Berkson & Gage, 1952; Cutler & Axtell, 1963; Haybittle, 1965) and the efficacy of an advertising program (Anscombe, 1961). In addition, Carr-Hill and Carr-Hill (1972) and Greenberg (1978) studied similar conceptual models of recidivism. And additional empirical justification comes from a recent (December, 1979) six-year follow-up study of offenders released from British prisons (Philpotts & Lancucki, 1979). Their data (see Figures 7-7 to 7-9) also strongly suggest the split-population model.

This model achieves more than just an improved fit to data; it suggests a new definition of recidivism. Using the difference of proportions test, the best program is the one that has the lowest proportion of recidivists at the end of, say, one year. Using the exponential distribution, the best program is the one whose participants fail at the slowest rate. Using the split-population distribution, the best program is the one which results in fewer participants ultimately recidivating. This last definition of "best" is intuitively more appealing than the other two, and is proposed here as a new operational definition in a correctional context.

A Behavioral Explanation of the Split-Population Distribution

The split-population model can be generated by a related model we

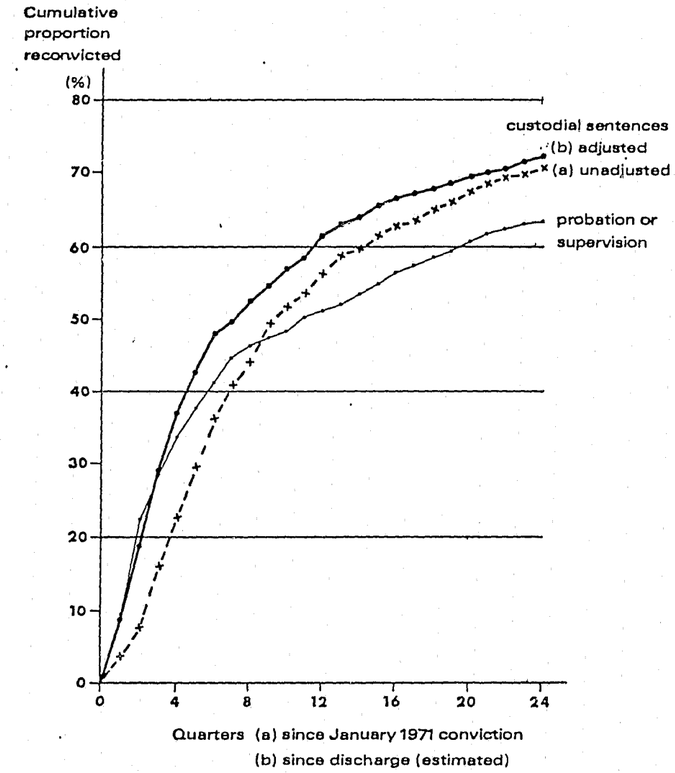
Males convicted of standard list offences in January 1971: cumulative proportion reconvicted by sentence and time since January 1971 conviction



Source: Philpotts and Lancucki, 1979

Figure 7-7 Reconviction Rates for Great Britain, by Type of Penalty

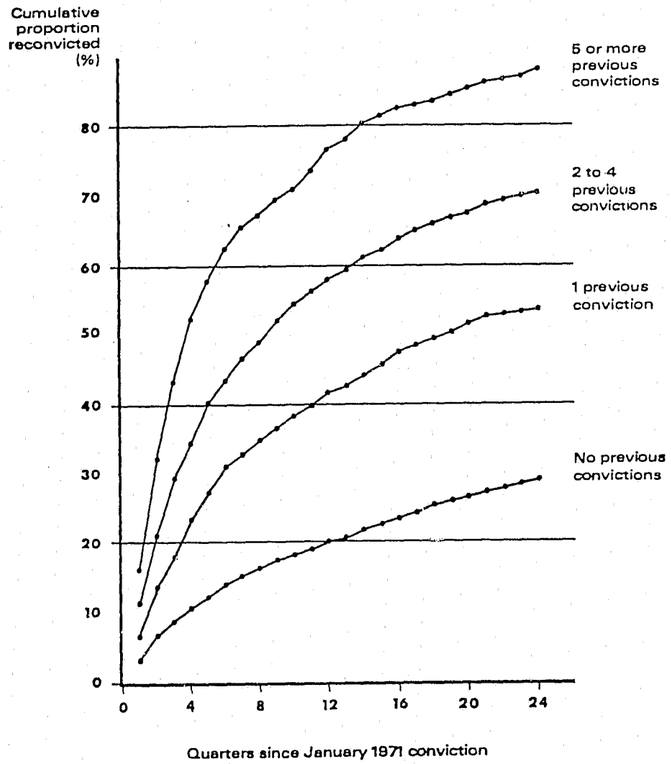
Males convicted of standard list offences in January 1971 and given custodial sentences: cumulative proportion reconvicted by (a) time since January 1971 conviction and (b) time since discharge (estimated)



Source: Philpotts and Lancucki, 1979

Figure 7-8 Reconviction Rates for Great Britain, Adjusted for Sentence Length

Males convicted of standard list offences in January 1971: cumulative proportion reconvicted by number of previous convictions and time since January 1971 conviction



Source: Philpotts and Lancucki, 1979

Figure 7-9 Reconviction Rates for Great Britain, by Number of Previous Convictions

have called the "critical time" model. It assumes that an individual has a constant failure rate (i.e., as in the exponential distribution) for a length of time. This length of time is the person's criticial time: if he "stays clean" for that length of time he then can be considered rehabilitated -- a success -- with essentially zero failure rate thereafter. The critical time may not be the same for all individuals; some may essentially be rehabilitated quickly while others may require a good deal of time outside before they can be termed successes. It can be shown (see Appendix A) that if the critical time is exponentially distributed over the population, then the failure times have the same distribution as in the split-population distribution. In other words, the split-population model can be reinterpreted as the critical time model.

Bloom (1978) has proposed a model quite similar (Kaplan, 1979) to the critical time model; however, conceptual and computational problems with his model limit its usefulness.

The Mixed Exponential Distribution

A three-parameter distribution that has been investigated recently (Harris, Kaylan, & Maltz, 1980) involves a mixture of two exponential distributions. It is a logical extension of the split-population distribution. This type of distribution has also been suggested by Carr-Hill and Carr-Hill (1972) and by Greenberg (1978).

The mixed exponential distribution also posits that two subgroups will be generated. The fact that distinguishes it from the split-population distribution is that both subgroups fail -- but at different failure rates. One failure rate is considerably larger than the other; this rate can be considered the rate at which the eventual failures fail. The lower failure rate can be considered the "ambient" failure rate. This is the rate at which the population in general might be expected to commit offenses. (Figure 7-10 depicts this distribution).

This extension recognizes that a failure rate of zero for the "successful" subgroup may be incorrect. It is also reasonable to assume that there is a (comparatively) low failure rate for individuals in that subgroup, equivalent in magnitude to the probability that an ordinary citizen fails (i.e., is arrested for a crime).

Model Selection

We see, then, that a number of different models of recidivism have been spawned by the original Stollmack and Harris (1974) paper. One of the difficulties that remains, however, is selecting the most appropriate model from among these (and other) choices.

Little guidance is offered in the literature on how to select the most appropriate model from among the candidates. Box and Tiao (1973: 8), for example, discuss this problem (they call it "model criti-

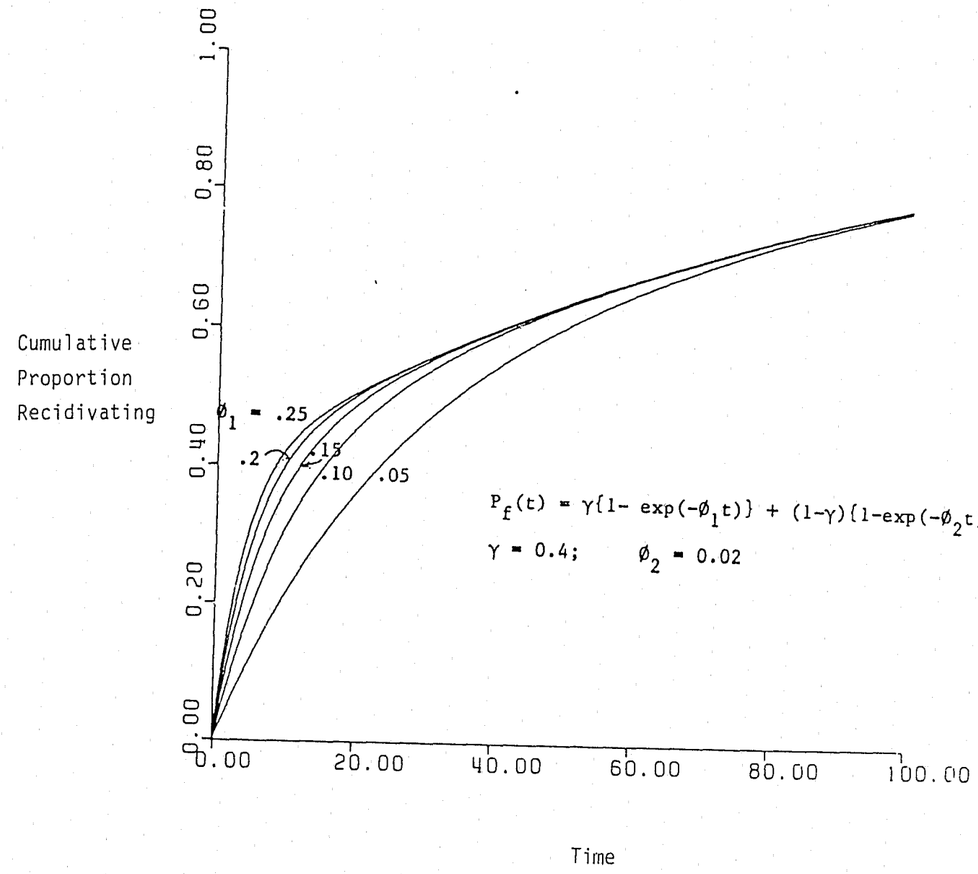


Figure 7-10 The Mixed Exponential Distribution

cism"), and suggest using the chi-square goodness of fit test. In most other texts as well the problem of model selection is equated with the problem of goodness of fit, i.e., with how well the model fits the data.

The chi-square goodness of fit test is the one most commonly employed for this purpose (e.g., Carr-Hill & Carr-Hill, 1972; Witte & Schmidt, 1977; Harris et al, 1980). The chi-square statistic is a measure of the discrepancy between the empirical data and that produced by the model's estimates. It is very useful, especially for those applications in which the desired goal is fitting a curve to data (e.g., Daniel & Wood, 1971).

But the model with the best statistical fit is not necessarily the best model. A model may fit the data very well within the region for which data are available, but it can exhibit unreasonable behavior outside this region. In selecting a model one should also consider the way the model is to be used and the nature of the process under study. If the model is to be used for extrapolation, it should be tested for its ability to extrapolate correctly. Thus, in Chapter 9, we offer empirical evidence suggesting that models of the recidivism process based on the split-population and mixed exponential distributions are more appropriate than other models, in terms of their ability to extrapolate beyond the available data. We can also justify this choice on theoretical grounds, based on the nature of the crim-

inal justice process and on the statistical properties of thinned and superimposed point processes.

Theoretical Considerations in Model Selection

It should first be recognized that we are not dealing with every possible failure event (i.e., violation of parole conditions or of the criminal law) that occurs. Not every violation that occurs will be observed by the relevant authorities. In statistical terms, if we consider the occurrence of violations by an individual to be a point process generated by some distribution, then the observed violations are a "thinned" point process, thinned because not all of the failure events are observed (Figure 7-11). Haight (1967: 25) refers to a finding by Renyi con.erning thinned point processes: regardless of the nature of the violation-generating process, if the thinning process is independent of it (i.e., not linked to the offender) and if most of the points are thinned, then the resultant thinned process is Poisson.

These two conditions hold in the case of recidivism. The violation process can be considered independent of the arrest process, since the former is a personal characteristic of the recidivist and the latter is a characteristic of the agency response. It also appears to be true that most of the points are thinned -- that is, most of the violations go undetected: if the clearance rate for felonies

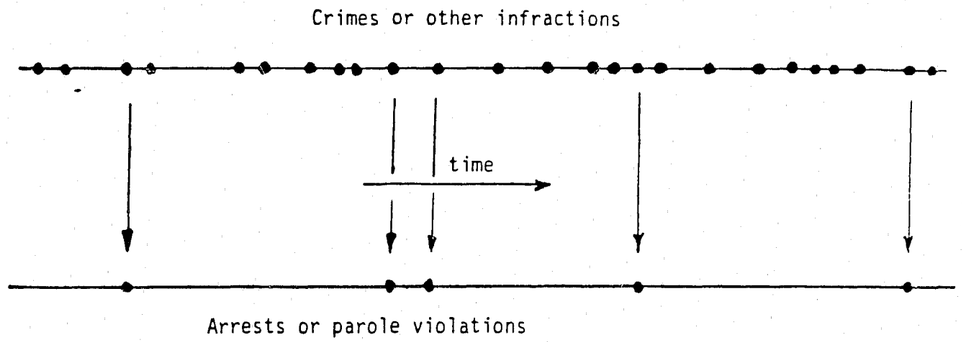


Figure 7-11 The "Thinning" of One Process (Crimes) by Another (Arrests)

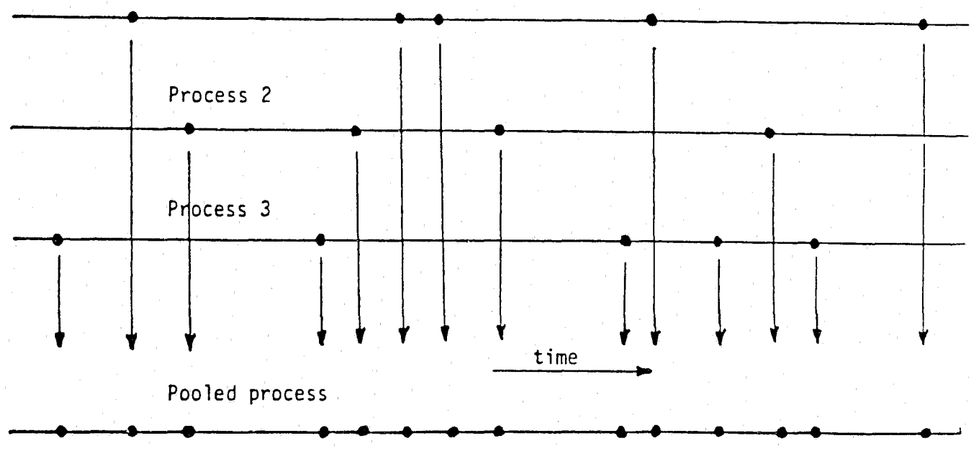


Figure 7-12 The Superposition of Individual Processes to Form a Pooled Process

is about 20 percent (see Boland & Wilson, 1978), then for misdemeanors and parole violations it must be much lower. Therefore, one would expect the Poisson distribution to typify the distribution of times between a recidivist's observed violations, his thinned point process.

The thinning process refers to an individual's event distribution: we must also consider the statistical properties of the cohort. To do this we consider the point process formed by superimposing the individual (observed) point processes, to obtain a pooled process (Figure 7-12). If we assume that the individuals in the cohort act independently of each other, arguments similar to those used by Cox and Miller (1965: 362-364) can be used to show that the pooled process will approach a Poisson distribution -- again, regardless of the nature of the individual point processes.

The Poisson distribution is a limiting distribution in the study of point processes, applicable when studying the superposition or thinning of independent processes. In a Poisson process the intervals between event occurrences are exponentially distributed, as is the time to occurrence of the first event. Since it is this (first) time interval with which we are concerned, the assumption of an exponential distribution for recidivism events thus appears to be strongly supported on theoretical grounds.

NOTES

1. In Chapter 9 we analyze Uniform Parole Reporting data from 32 states and point out the difficulties of comparing the results.

2. One way to insure that the two groups are similar is to use matched pairs and to assign individuals from each pair randomly to experimental or control groups. But there are two difficulties with this procedure: First, the characteristics that should be matched are open to question --prior record, of course, but what about height and weight, IQ, vocational skills? Second, the more factors to be matched (and the more levels for each factor), the fewer the number of individuals that can be matched.

3. An interesting case in point is the study of preventive police patrol conducted in Kansas City, Missouri (Kelling et al., 1974). The areas which were to receive twice the patrol coverage were called proactive patrol areas. These areas were matched with control (normal patrol) areas and with areas which were to receive no preventive patrol coverage, termed reactive patrol areas. The latter term was used to disguise the fact that some areas were to receive no preventive patrol coverage, so that citizens would not get up in arms about

their ostensible loss of protection. Although experimental conditions were not maintained completely (Larson, 1975), the deception nevertheless worked.

4. One well-known exception to this is the Provo Experiment (Empey and Erickson, 1972).

5. In such a case a favorable outcome for the program participants may only mean that the judge is proficient at picking the best risks for program participation. Of course, if the program outcome is unfavorable one can make the opposite argument.

6. The ethical considerations described above are part of the argument that "we should not experiment with human lives". However, the other side of the coin should be looked as well -- we continue to conduct uncontrolled experiments with human lives, although they are not described as experiments: every time a new law is passed that specifies sentences for offenses, or elibibility requirements for welfare funds or medicaid, it is based on a belief that it is an improvement over the present situation. One then must address the question: is it more ethical to impose new policies without attempting to evaluate their effectiveness before they are adopted wholesale, or is it more ethical to attempt to evaluate the policies on a small-

scale experimental basis first? As Mosteller points out (Sechrest et al, 1979: 71), "The only alternative to experimenting with people is to fool around with people."

7. The ethical considerations involved in measuring success or extent of success were given in Chapter 3. Practical considerations also dictate against the use of multi-valued measures of success or failure. Data on the status of releasees tend to be quite unreliable. One parole officer may rely on a parolee's word concerning employment, a second may talk to his employer, a third may visit the parolee in his workplace. In some cases revocation for a technical violation may mask a new offense, in others it may reflect animus between parole officer and parolee. With data of this caliber it is pointless to try to make fine distinctions. Therefore, a binary measure of success (i.e, success/failure) can be just as appropriate as a continuous measure for recidivism studies.

Furthermore, parsimony and simplicity are also desirable in developing a model -- so long as the salient characteristics of the process are not lost by oversimplification. Indeed, the addition of complexity to the model may make it more difficult to understand the nature of the process.

8. An exception to this is the study by Kitchener et al. (1977). They report on an eighteen-year follow-up of a cohort of individuals released from federal prison. The number of failures kept increasing throughout the eighteen years (see Figure 7-5), but at a negligible rate after ten years.

9. Programs such as HUD's Model Cities and LEAA's Impact Cities were attempts to take pilot programs and scale them up to cover an entire city or section of a city. The record of these programs, to say the least, was spotty. In the education field, programs such as the "new math" and innovative reading programs are so often marked by their success as experiments and their failure as curriculums. See Levitan & Taggart (1976) for a critique of these programs.

10. The residents of this study were lodged in halfway houses funded by the Minnesota Governor's Commission. "Unsatisfactory" clients were those who did not complete their residence satisfactorily, for reasons which included revocation or conviction for additional offenses while in residence as well as other non-criminal reasons.

11. Figure 7-1 leads one to make some interesting conjectures. Perhaps "satisfactory" clients are typified by the following characteristics: they had high expectations for the program's ability to

help them, expectations that were not met by the program; or they were con men who just went along with the program to escape a less desirable alternative; or they were the type who always exhibit compliant behavior and are easily swayed, complying with the expectations of program personnel when in residence and with the expectations of their peers when released. Similarly, the negligible amount of recidivism after six months, shown by the "unsatisfactory" clients who were out longer than six months suggests that they may be the ones with more backbone, who are not easily swayed, and who did not expect miracles from the program. Looked at in another way, perhaps they are the program's successes.

12. For example, if ten people fail in the time interval between the 36th and 37th week, and there were one hundred people still (unfailed) in the program at the start of the 36th week, then the failure rate at week 36 is 10/100 = 0.1/week. If nine people subsequently fail during week 37, the number of people remaining at the start of week 37 is 100-10 = 90, so the failure rate is 9/90 = 0.1/week again. As can be seen from this example, a constant failure rate does not mean that the same number of people fail in each time interval: it means that the population declines by the same proportion in each interval.

13. The failure rate need not be constant. For example, electromechanical components are sometimes characterized by a "bathtub"-shaped failure rate (Barlow & Proschan, 1975: 55); i.e., a high initial failure rate declining to a lower failure rate in middle periods and increasing again in later time periods. The initial high failure rate is usually attributed to defects in manufacture ("lemons") and the later high failure rate to the components' wearing out.

14. Under some conditions these kind of distributions do not have certain well-behaved properties; consequently, they are known (Feller, 1966: 127) as "defective" or incomplete distributions. We hasten to point out that models based on these distributions are not flawed in any way, but that their statistical properties are more complicated.

15. There appears to be an exception to this general rule. In a study of automobile scrappage rates, Golomb (1978) found that failure of an automobile is not inevitable. If a car has survived to a certain age in good condition, it is apparently worth more to the owner to keep it up (as an "antique" vehicle) than to scrap it. Although the individual components in the automobile may fail, the automobile does not.

16. Some of the non-failures are not true successes. We are dealing with reported failures and not all recidivism events are reported. In addition, people in this category may have failed, but in another jurisdiction which does not forward reports to the jurisdiction doing the study. But absent information to the contrary, we will call members of this subgroup "successes".

17. The failure times of this subgroup, then, are assumed to be exponentially distributed. From the standpoint of the whole group the failure rate is not constant but decreases over time.

18. Partanen (1969) also suggested a model of this type, as one of a family of decreasing failure rate models.

19. I am indebted to Michael A. Greene of American University for pointing this out to me.

CHAPTER 8

PARAMETER ESTIMATION

In the preceding chapter we described a number of models of the recidivism process, and explained why we focus on the split population and mixed exponential models. In this chapter we describe a maximum likelihood technique which we use to estimate the parameters of the split-population model. Based on the model we are able to write expressions for the probability of occurrence of certain events. These probability expressions are then used in the development of a function, called the likelihood function, that incorporates the empirical data. Using the likelihood function we can develop estimates of the model parameters.

From Model to Probability Expressions

In this section we will use a simple but useful version of the split population model for illustrative purposes. We first assume that the members of the population under study have all been released from the program simultaneously.[1] Thus, when the failure data are obtained all of the members of the population will have had the same exposure time since release -- with the exception, of course, of those who have failed in the interim and have been returned to prison.

The model is based on the premise that two subpopulations will eventuate, one (of fraction 1-γ) consisting of those who will ultimately not fail, the other (of fraction γ) consisting of those who will eventually fail. We assume that a fraction p of the to-be failures fails each month,[2] and the remaining fraction q of the to-be failures survives to the next month, where

$$q = 1 - p \qquad (1)$$

To make the analysis concrete, let us assume that the population under study consists of 400 persons, and that 25 percent of them (or 100) will eventually fail. In other words, N = 400, γ = 0.25, and $N\gamma$ = 100. Then $N(1-\gamma)$ = 300 will not fail. Let the fraction p of eventual failures who fail in any given month be 0.1. Then during the first month ten percent, or exactly ten, of the to-be failures will fail. This leaves 390 persons at the start of Month 2: the 300 who will never fail and 90 (= $N\gamma q$) who will eventually fail but have not done so yet.

During the second month another ten percent of the to-be failures fail. Since there are now 90, nine persons will fail during Month 2, leaving 81 (= $N\gamma q^2$) who will eventually fail but have not done so yet.[3] And so on.

Therefore, the total number of people failing by a given month m would be given by

$$N_f(m) = N\gamma(1-q^m) \qquad (2)$$

where the subscript f signifies "failure." For our example, with m = 2, we obtain

$$N_f(2) = 400 \times 0.25 \times (1-0.9^2) = 19 \text{ people,}$$

as we noted before.

From Equation 2 we see that the probability of failing by month m (called the cumulative distribution of failure times) is given by

$$P_f(m) = \gamma(1-q^m) \qquad (3)$$

For example, if γ were 0.25 and q were 0.9 (as in the numerical example above), Figure 8-1 would depict the cumulative distribution of failure times.

Equation 3 and Figure 8-1 represent the discrete version of the continuous split-population distribution shown in Figure 7-6. This representation is often more useful than its continuous counterpart because of the way recidivism data are usually collected. Failure

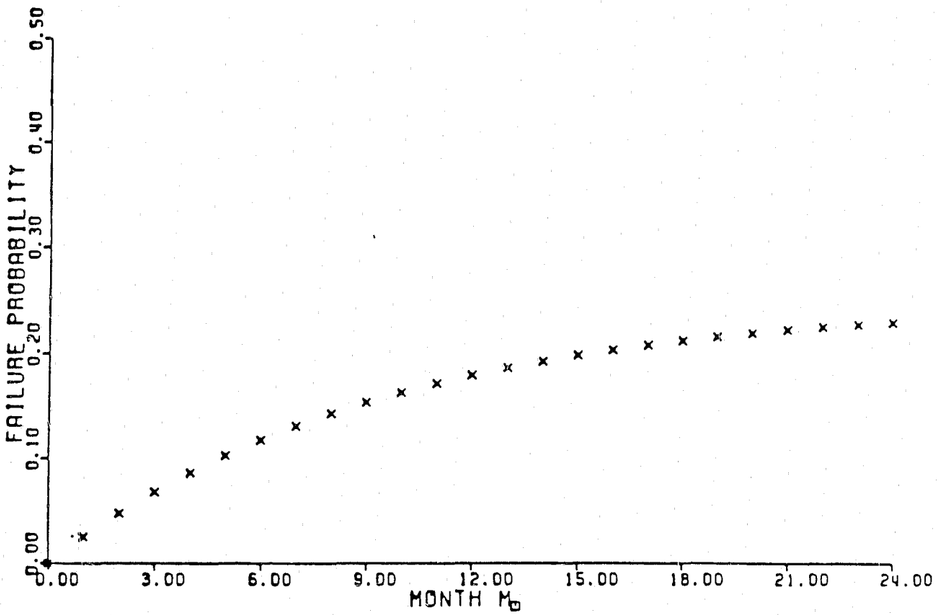


Figure 8-1 Cumulative Distribution of Failures for $\gamma = 0.25$ and $q = 0.9$

data are often presented in tables listing the number of individuals who failed in the first month after release, the number who failed in the second month, etc., rather than the exact date of failure for each. Since we do not know when within the month each person failed, a discrete distribution would be more appropriate than a continuous one.

The number of people who fail <u>in</u> month m can be computed from Equation 2. It is equal to the number of people who failed <u>by</u> month m, less the number who failed <u>by</u> month m-1, or

$$N_f(m) = N\gamma\,(1-q^{m}) - N\gamma\,(1-q^{m-1}) = N\gamma\,q^{m-1}(1-q) \qquad (4)$$

and the probability of failing in month m is therefore

$$p_f(m) = \gamma q^{m-1}(1-q) \qquad (5)$$

Some people have not failed by month m, and we never find out whether or not they do fail. This may be because we lose track of them (e.g., they are discharged from parole) or because we have fixed a date to stop data collection and they have not failed by then. But they have been exposed for m months without failing, either because they will never fail or because they eventually fail but have not yet

done so. The number of such people is (the subscript s signifies "success")

$$N_s(m) = N(1-\gamma) + N\gamma q^m \tag{6}$$

The first term in Equation 6 represents those who will never fail. The second term represents those who will fail eventually, but have not done so yet. [Note that when we add $N_s(m)$ to $N_f(m)$ (from Equation 2) we obtain N, as we would expect, since a person either has or has not failed by month m.] And as before, the probability of not failing by month m is

$$P_s(m) = 1 - \gamma + \gamma q^m \tag{7}$$

Using Equations 5 and 7 we can now start the process of developing estimates of the parameters γ and q. The next step uses these equations to incorporate the data into the model.

From Probabilities to Likelihood Function

We now develop the likelihood function from Equations 5 and 7 with the help of another assumption. We will assume that every individual in the population under study acts independently of every other individual, that the success or failure of one person has no influence on any other person.

For illustrative purposes, let our population consist of ten people, of whom three failed in month i, two failed in month j, and one failed in month k, while the remaining four had not failed through month n. The probability of obtaining these results can be obtained:

Since we have assumed independence, the probability of a joint event is just the product of the probabilities of the individual events. Then, from (5) and (7), we have

$$P[\{t\};\gamma,q] = \begin{matrix} & 3 & 2 & 1 & 4 \\ (& [p(i)] & [p(j)] & [p(k)] & [P(n)] \\ (& f & f & f & s \\ (& & & & \\ (& & & & \\ (& i-1 & 3 \quad j-1 & 2 \quad k-1 & 1 \quad n \; 4 \\ (& [\gamma q & (1-q)] \; [\gamma q & (1-q)] \; [\gamma q & (1-q)] \; [1-\gamma+\gamma q \;] \end{matrix}$$

where $\{t\}$ represents the times $\{i,i,i,j,j,k;n,n,n,n\}$.[4] By collecting terms we can simplify this expression:

$$P[\{t\};\gamma,q] = [\gamma(1-q)]^{K} \, q^{T-K} \, [1-\gamma+ q \;]^{n \; 4} \tag{8}$$

where $T = 3i + 2j + k$ and $K = 3 + 2 + 1 = 6$ (the number of failures). Equation 8 is the likelihood function of the model described by Equation 2 and containing the data $\{t\}$. It may written as $L(\gamma,q)$.

In general, with a population of size N we would have k_1 persons failing within Month 1, k_2 within Month 2, . . ., and k_n within Month n; and M persons observed for n months without failing. For this

situation the likelihood function is

$$L(\gamma,q) = [\gamma(1-q)]^{K} q^{T-K} [1-\gamma+\gamma q]^{n} \qquad (9)$$

where $K = \Sigma_i k_i$, $B = \Sigma_i i k_i$, and $M = N - K$.

Equation 9 represents the probability of obtaining the given data {t} when Equation 2 represents the recidivism process. Since we have the data we can substitute specific values of γ and q into (8) and find the probability of obtaining these data. If we change the values of γ and q we can find the probability of obtaining the same data with the new values. In fact, we can vary γ and q throughout their entire ranges (from 0 to 1 for both variables) and, for example, find the values of γ and q for which the probability is highest, that is, for which the likelihood function is a maximum. These values of γ and q are called maximum likelihood estimates.

One other point is worth noting from Equation 9. For this model all of the data are summarized into four statistics: K, the number of failures in the population under study; N, the total number in the population; T, the total number of months the failures were exposed before failing; and n, the number of months the (as yet) non-failures were out and observed. If these four statistics are given, the likelihood function (9) is completely specified.

Characteristics of the Likelihood Function

In the numerical example given above, let us be more specific and set i = 3 months, j = 4 months, k = 5 months and n = 24 months. That is, three people fail in the third month, two in the fourth month, one in the fifth month, and the remaining four do not fail during the first two years of observation. In this case K, the total number of failures, is 6; N, the total population under study, is 10; T, the total failure time, is 22; and n, the observation time for the (as yet) non-failures, is 24. The likelihood function is thus given by

$$L(\gamma,q) = [\gamma(1-q)]^{6} q^{16} (1-\gamma+\gamma q^{24})^{4}$$

This equation is plotted against γ and q in Figure 8-2. It appears to be bell-shaped, so it can be aproximated by a bivariate normal probability density function. This property is of great utility for statistical purposes because of the important role the normal distribution plays in statistics. However, not all of the likelihood functions generated by our model of the recidivism process (3) have this property.

Figures 8-3 and 8-4 depict likelihood functions for two other sets of data. As in Figure 8-2 K = 6, N = 10, and T = 16. But in Figure 8-3 we used n = 18 months and in Figure 8-4, n = 12 months.

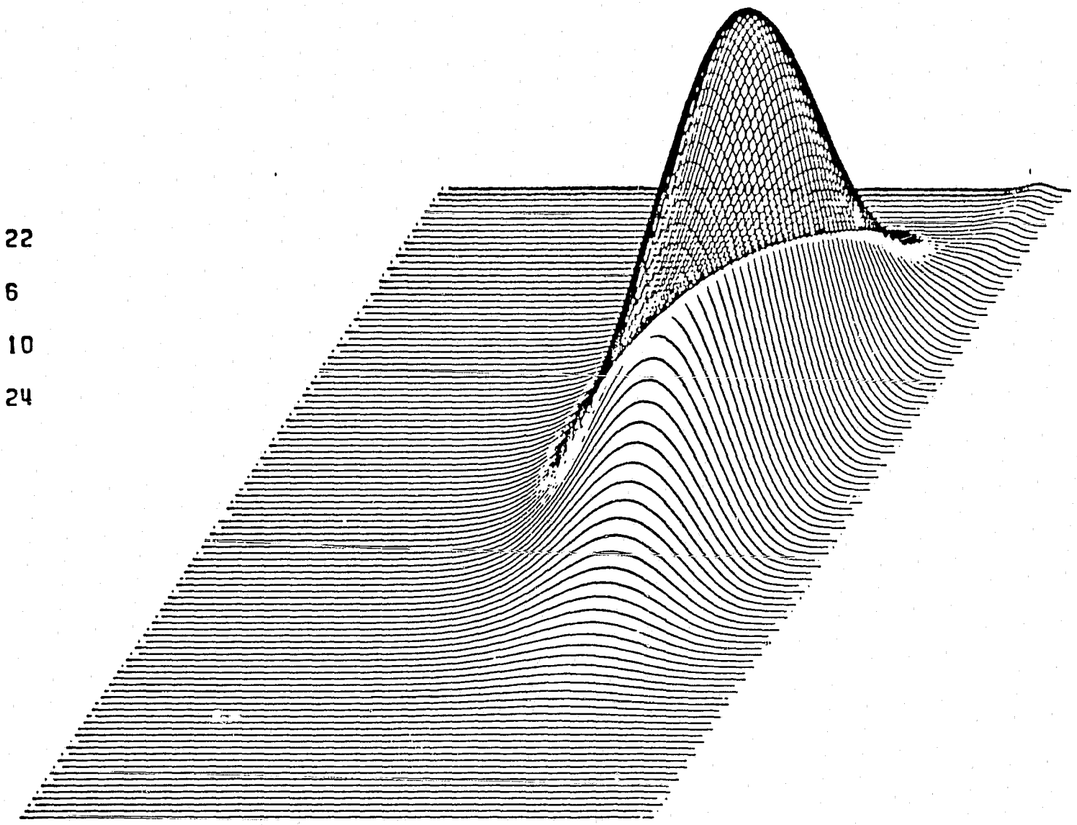


Figure 8-2 Likelihood Function for T = 22, K = 6, N = 10, with Observation Time n = 24 months

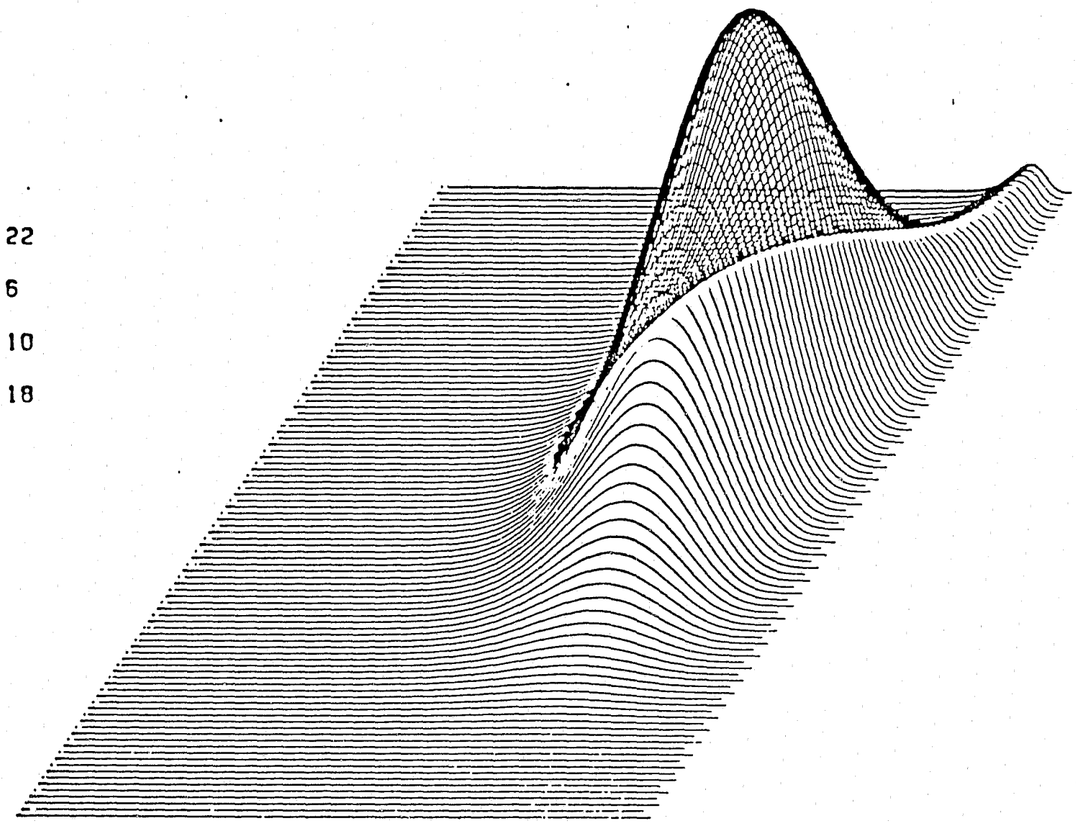


Figure 8-3 Likelihood Function for T = 22, K = 6, N = 10, with Observation Time n = 18 months

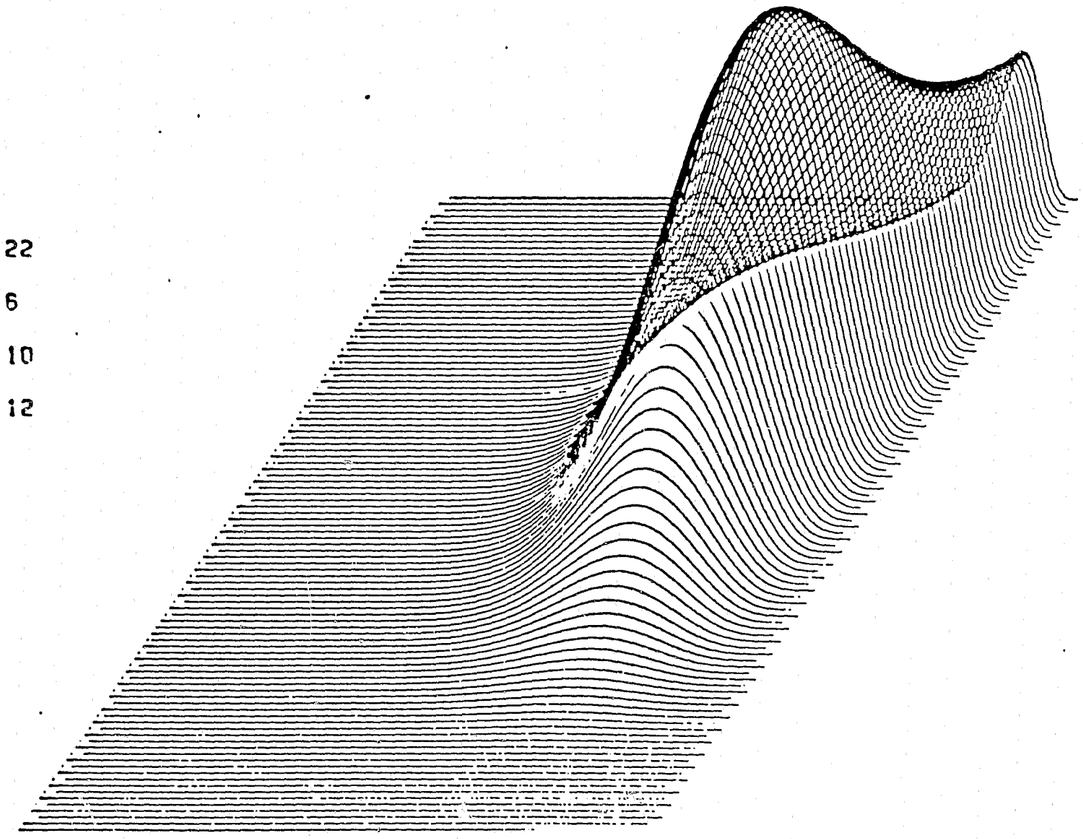


Figure 8-4 Likelihood Function for T = 22, K = 6, N = 10, with Observation Time n = 12 months

Figure 8-2 can be approximated quite closely by a two-dimensional normal probability density function. This is generally true for likelihood functions that meet certain regularity conditions (Rohatgi, 1975: 384). And this approximation is very useful for making confidence statements about parameter estimates or for testing hypotheses, since the normal distribution is tabulated and available to researchers. In fact, Lloyd and Joe (1979) used this approximation to make confidence statements about their estimates and to test hypotheses.

Unfortunately, the probability distribution (3) does not always meet the regularity conditions.[5] This does not necessarily mean that all likelihood functions based on this distribution cannot be approximated by a two-dimensional normal distribution. In fact, Figure 8-5 shows the magnitude of the error associated with this approximation for the likelihood function of Figure 8-2. As can be seen, the error is quite small.

The errors are considerably more significant if a normal approximation is used for the likelihood functions in Figures 8-3 and 8-4. The errors associated with such approximations are shown in Figures 8-6 and 8-7, respectively. This points out the need for care in making the assumption that the likelihood function is asymptotically normal.

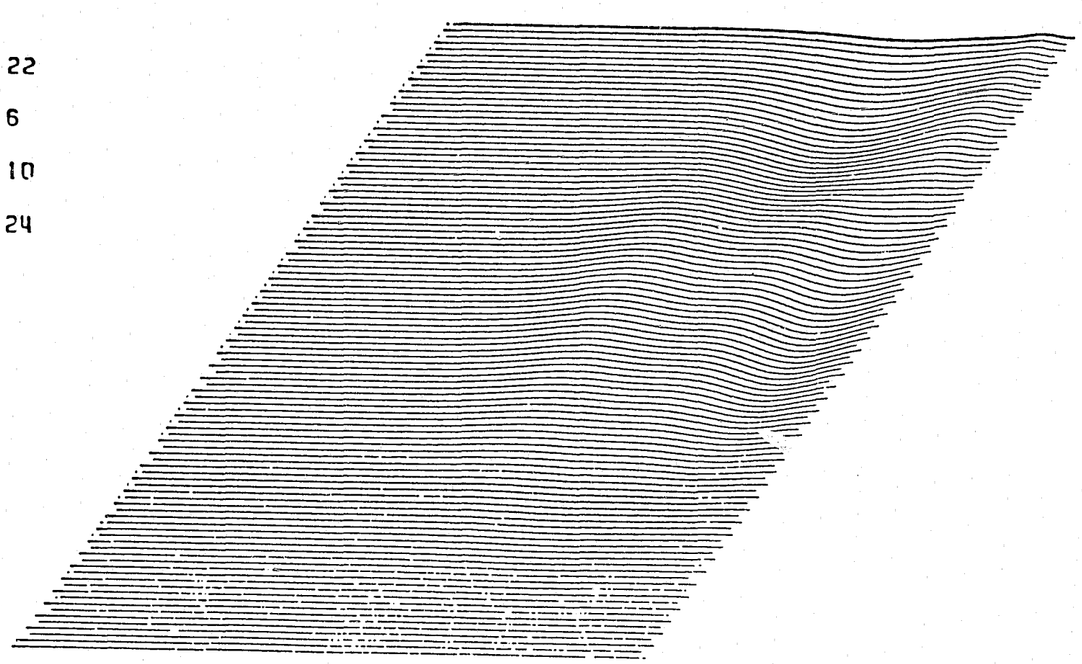


Figure 8-5 Error Associated with Approximating Figure 8-2 by a Normal Distribution

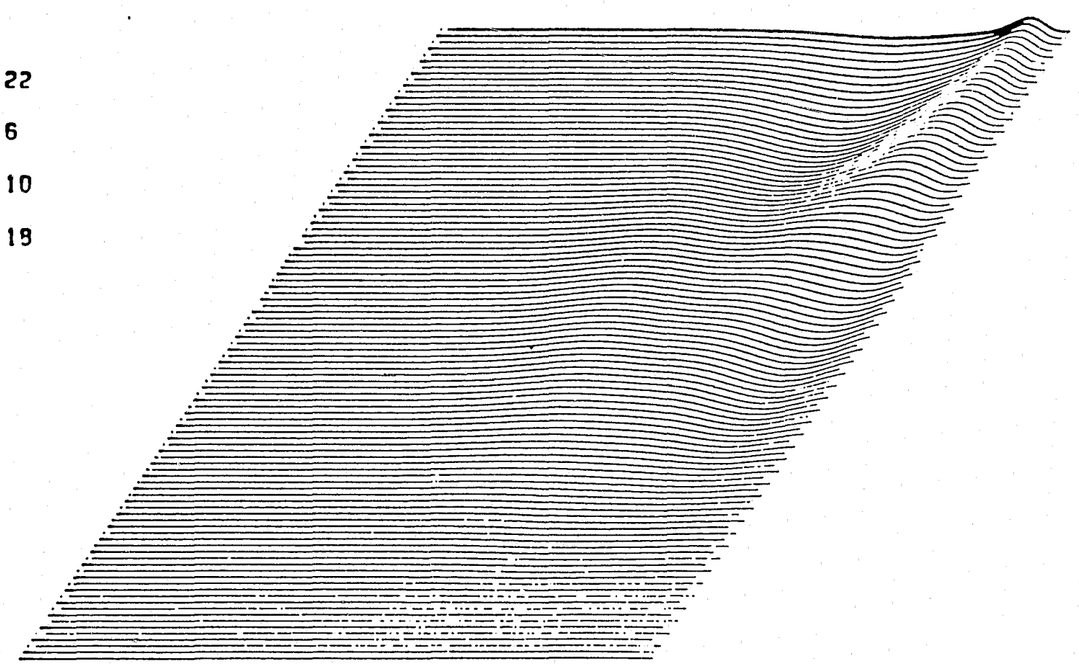


Figure 8-6 Error Associated with Approximating Figure 8-3 by a Normal Distribution

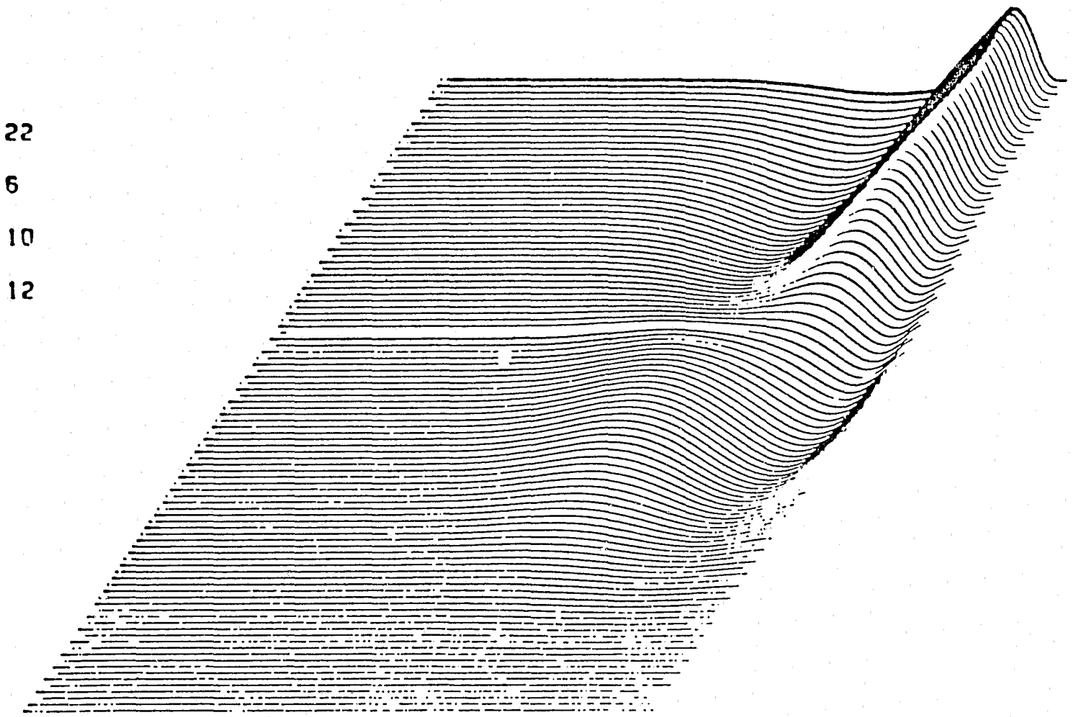


Figure 8-7 Error Associated with Approximating Figure 8-4 by a Normal Distribution

From Likelihood Function to Parameter Estimation

When the likelihood function is unimodal and symmetric, as in Figure 8-2, one selects the best estimates for the parameters γ and q by choosing the values at the modal value of the likelihood function. In such a case the three measures of central tendency -- mean, median, and mode -- give the same results.

When we have an asymmetric likelihood function as in Figure 8-4, however, the choice of best estimates is not quite so obvious. The values for γ and q at the mean, median, and mode of the likelihood function are all different. We therefore must be more specific about what we mean by the "best" estimate.

One can, for example, define "best" as "most likely". That is, we can choose for our estimates the values of γ and q at the maximum value of the likelihood function, as before -- hence, the term "maximum likelihod estimation". Conceptually, this estimation procedure is quite straightforward. Any one of a variety of hill-climbing techniques can be used to find the values of γ and q for which the likelihood function is a maximum.

But the maximum likelihood estimation procedure is not always indicated. Its use implies that the likelihood function has a local maximum within the feasible region (i.e., $0 \leq \gamma \leq 1$, $0 < q < 1$). As we see from Figure 8-8, this is not always the case.[6] The likelihood function in Figure 8-8 does have a maximum in the feasible region, but

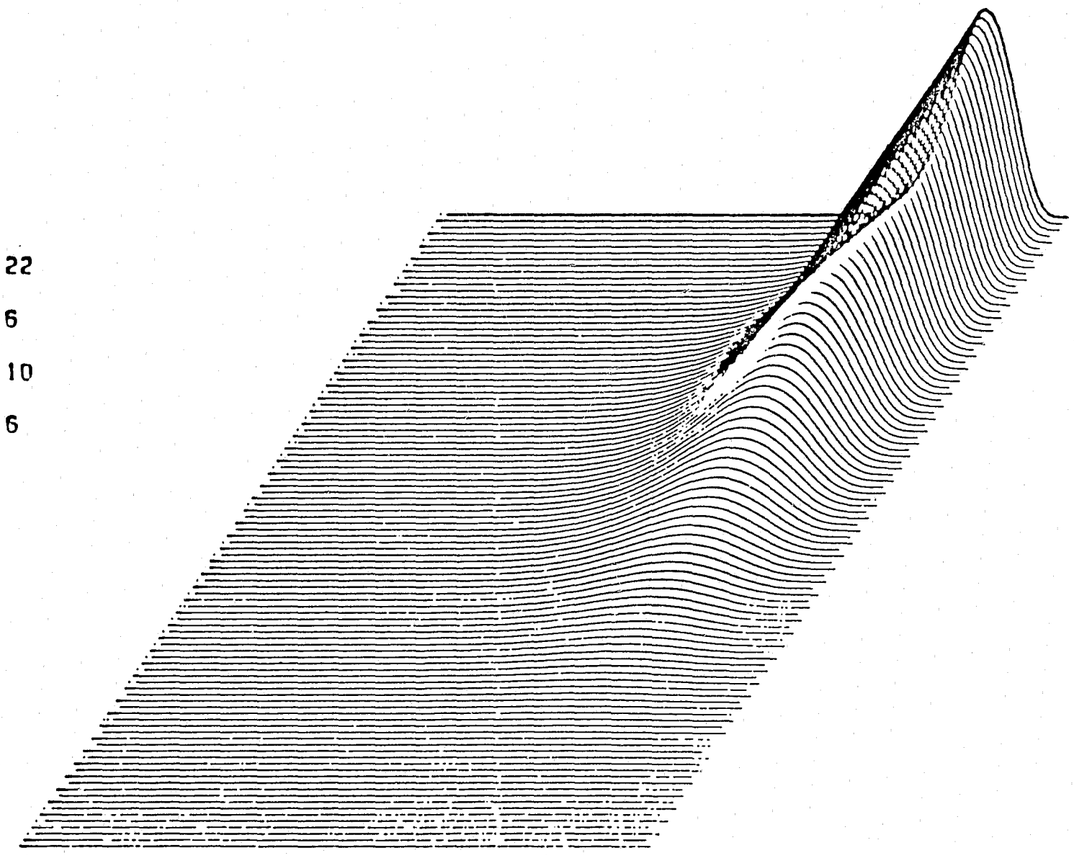


Figure 8-8 Likelihood Function for T = 22, K = 6, N = 10, with Observation Time n = 6 months

it is at the region's border, at $\gamma = 1$. Since this is not a local maximum, it may not be the best estimate of γ.

For this case a more suitable measure of central tendency might be the mean. A Bayesian estimation technique can be used to calculate the mean values of γ and q; see Maltz and Pollock (1981). Although this technique requires the specifications of the a priori probability distributions of γ and q, the results are normally fairly stable over a wide range of prior distributions.

A likelihood function can have a local maximum or it can have no local maximum, as discussed above. But it can also have more than one local maximum. In fact, bimodal likelihood functions crop up in cases of interest to us. Recall that we have discussed two related models: the split population model, in which only a fraction γ of the population fails; and the mixed exponential model, in which all fail, but the population fraction fails at a more rapid rate than the population fraction $1-\gamma$. The cumulative distribution for the mixed exponential model is[7]

$$F(j) = \gamma\,(1-q_1^{\,j}) + (1-\gamma)(1-q_2^{\,j}) \qquad (10)$$

To show how a bimodal likelihood function results, let us assume that a local maximum of the likelihood function occurs at $\gamma = 0.42$, $q_1 = 0.80$ and $q_2 = 0.99$. The symmetry of Equation 10 dictates that

another local maximum (of equal magnitude) must occur at $\gamma = 0.58$, $q_1 = 0.99$ and $q_2 = 0.80$.

A symmetrical bimodal likelihood function poses no major difficulties. One merely looks at the region in which $q_1 < q_2$, which when added to the other constraints on γ, q_1 and q_2, further delimits the feasible region. However, for a mixture of two different distributions -- say, an exponential with a Weibull or lognormal or other distribution -- there would not be the same symmetry and other procedures would have to be employed. Fortunately, we do not have to worry about this added complication.

NOTES

1. Although this assumption is not necessary for the exposition, it allows a reduction in the number of variables we have to consider.

2. For the sake of convenience, we use one month as the basic time interval in this chapter. Then the variable p is the probability that an individual at large fails in any given month, given that the individual will eventually fail.

3. Of course, we stop the numerical example here because in Month 3 we would find 8.1 persons failing, which would be difficult to achieve in real life. The purpose of the example is to give the reader an appreciation of the way the variables N, γ, p, and q interact, and should not be slavishly followed.

4. Depending upon the way the problem is structured, there may be a constant multiplying the likelihood function, equal to $\binom{N}{K}$, where N is the size of the population and K is the number of failures. See Stollmack (1979) and Maltz, McCleary and Pollock (1979) for a discussion of this point. However, whether or not the constant is included does not affect the results of this section.

5. One of the regularity conditions is that $F(\infty) = 1$. For our model of the process (3) we see that $F(\infty) = \gamma \leq 1$.

6. Figure 8-8 is the likelihood function for the same numerical example discussed previously, except that the non-failures are observed for only six months, so n = 6. Thus it is not a "pathological" case, but is entirely realistic.

7. We use the discrete version of the mixed exponential model here; the argument applies to the continuous model as well.

CHAPTER 9

APPLYING THE METHODS

In this chapter we report on applications of the methods described in Chapter 8 to data sets which had been analyzed previously using other techniques. The data are from studies undertaken by the Illinois and North Carolina departments of corrections, the United States Parole Commission, and the United States Bureau of Prisons. In addition, parole data were obtained for 32 states from the Uniform Parole Reporting Program. These applications show different aspects of the methods, and demonstrate their applicability and their limitations.

Comparing Models: Illinois Data

One concern of the analyst is to determine which model of the process appears to be the most appropriate. In the paper in which we introduced the split-population model in a correctional context (Maltz and McCleary, 1977), we illustrated this use by applying it to a cohort of 257 Illinois parolees. This cohort has since been analyzed using other models of the recidivism process. Harris and Moitra (1978) used a Weibull model. Harris, Kaylan and Maltz (1979) used a mixed exponential model, and Bloom (1978) used a more complex model.[1]

The models and parameters (found using maximum likelihood estimation) are given in Table 9-1. Figure 9-1 depicts the associate failure distribution curves these models generated, superimposed on the data. These models can be compared in two ways, graphically and using a standard goodness of fit test.

As can be seen from the figure, it would be difficult to select one model of the recidivism process over another in terms of its "eyeball" fit to the data. For example, the split population model and Bloom's model (Models 1 and 4) are essentially indistinguishable from one another. Both provide excellent fits to the data in the later months. The Weibull model (Model 3) provides a better fit than these in the earlier months but diverges strongly in the later months. The mixed exponential (Model 2) provides the best overall fit to the data, but does not fit as well in the later months as do the first two models.

"Eyeballing" to see which model fits best does have its advantages. One can take into account slight differences that more formal (quantitative) model selection techniques would miss. One can also include complex or subtle selection criteria that may defy quantification. Furthermore, it furnishes some insight into what one should use for criteria in developing formal techniques.[2] But by itself eyeballing is not really a suitable technique. The findings are not necessarily replicable over time, nor are they likely to be the same for

Table 9-1 Four Models of Recidivism Compared to Data for 257 Illinois Parolees

Model 1. **Split population exponential distribution:** $F(t) = \gamma(1-\exp(-\phi t))$

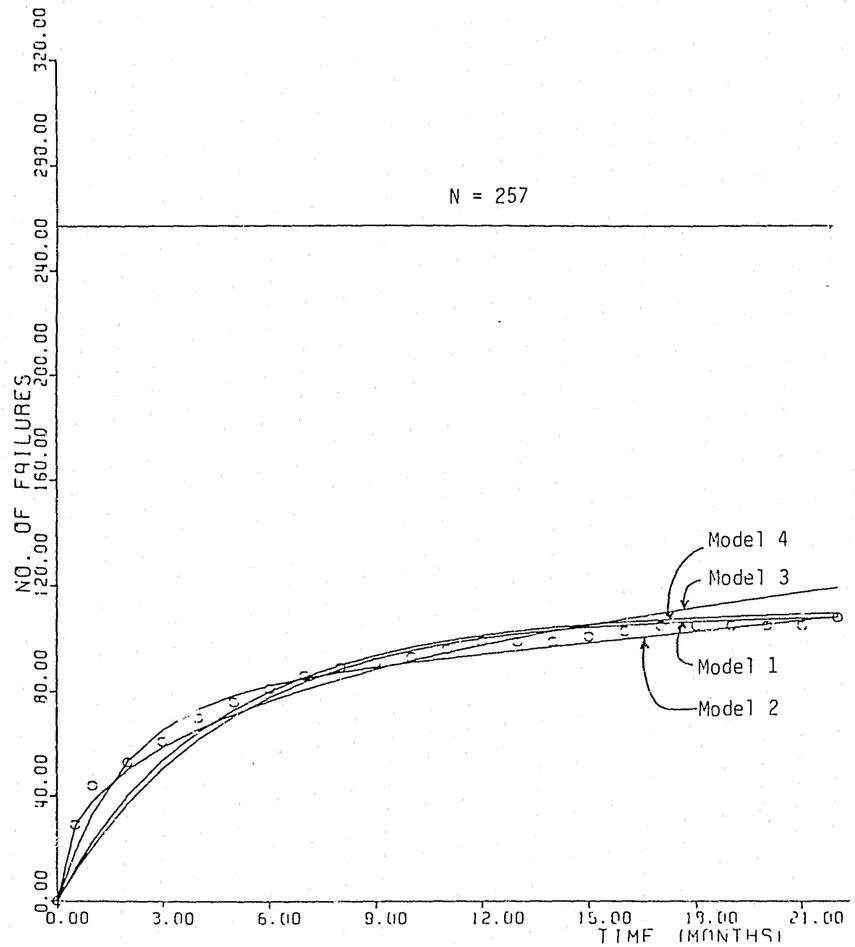
Model 2. **Mixed exponential dsitribution:** $F(t) = \gamma(1-\exp(-\phi_1 t)) + (1-\gamma)(1-\exp(-\phi_2 t))$

Model 3. **Weibull distribution:** $F(t) = 1-\exp(-(t/\eta)^{\beta})$

Model 4. **Double exponential distribution:** $F(t) = 1-\exp(-(b/c)(1-\exp(-ct)))$

Model	Parameter estimates
1	$\gamma = 0.4247$ $\phi = 0.2075$
2	$\gamma = 0.5411$ $\phi_1 = 0.2942$ $\phi_2 = 0.0090$
3	$\beta = 0.44$ $\eta = 64.39$
4	$b = 0.101$ $c = 0.178$

Time interval (months)	Actual number failing during that interval		Forecast number of failures			
			Model 1	Model 2	Model 3	Model 4
0 - 0.5	29		10.756	18.737	28.592	12.122
0.5 - 1	15		9.696	14.485	9.407	10.589
1 - 2	9		16.620	20.005	12.146	17.457
2 - 3	8		13.506	12.305	8.586	13.607
3 - 4	9		10.975	7.817	6.821	10.730
4 - 5	6		8.918	5.198	5.728	8.543
5 - 6	5		7.247	3.668	4.971	6.858
6 - 7	5		5.889	2.772	4.410	5.543
7 - 8	3	} 5	8.675	4.177	7.598	8.185
8 - 9	2					
9 -10	2	} 5	5.728	3.374	6.428	5.500
10 -11	3					
11 -12	3					
12 -13	0					
13 -14	0	} 5	6.280	5.995	10.548	6.310
14 -15	2					
15 -16	2					
16 -17	2					
17 -18	0					
18 -19	0	} 7	3.720	9.717	13.986	4.183
19 -20	0					
20 -21	0					
21 -22	3					
Over 22	149		148.988	148.751	137.779	147.373
Value of chi-square statistic			46.53	17.65	13.50	36.81
Degrees of freedom			10	9	10	10
Significance level			< 0.5	4.2	19.8	< 0.5



Model 1 - split population model
Model 2 - Mixed exponential model
Model 3 - Weibull model
Model 4 - Bloom model

Figure 9-1 Fitting Models to Illinois Data

different people. Furthermore, one cannot use it to determine how much better one model is than another.

For these reasons a quantitative measure, the chi-squared goodness of fit test, (e.g., Bury, 1975: 196) was applied to all four models. The results are given in Table 9-1. As can be seen, the Weibull model (Model 3) fits the data best according to this test. However, this may not be the appropriate test to use for models such as these.[3]

The purpose of these models is to estimate into the future, to project how many people will eventually fail over a period of time, and is not to achieve the best fit to the known data. For example, the dashed lines in Figure 9-2 represent the projection of these models out to 36 months. It would benefit the model selection process greatly to find out how many members of the cohort actually failed over the ensuing fourteen months, and thus have a measure of the models' relative forecasting abilities. Unfortunately, this information is not available.

Another strategy is to use only the first few months of data (e.g., 12) for the initial projection, then see how well the rest of the curve fits the data. Or alternatively, one can estimate the model parameters using varying amounts of data to see how stable the estimates are as more data are used for estimation. This was done using the split population model (Maltz and McCleary, 1977; Miley, 1979; Maltz and McCleary, 1979). Figures 9-3 and 9-4 show the estimates[4] of r and a based on two through 24 months of data.

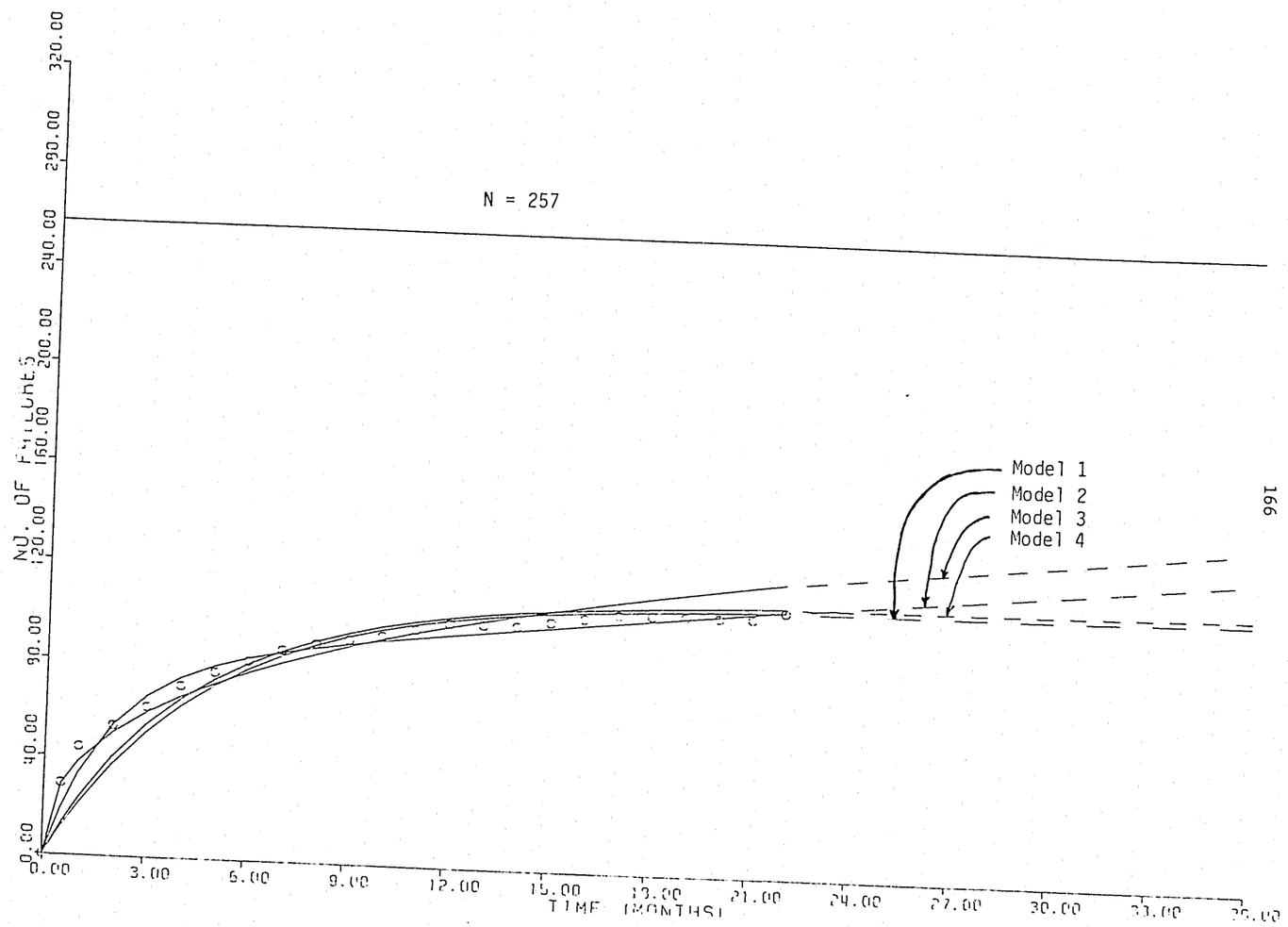


Figure 9-2 Extrapolating the Models Beyond the Data

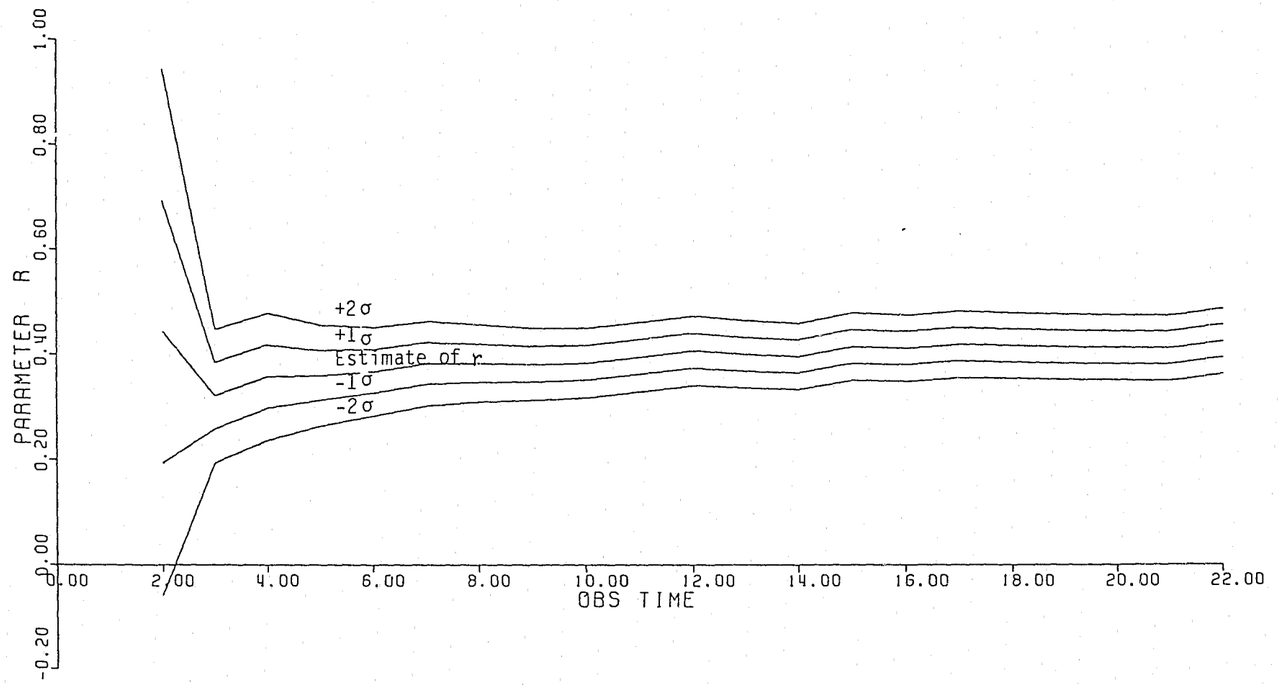


Figure 9-3 Variations in the Estimate of r as the Observation Time Increases

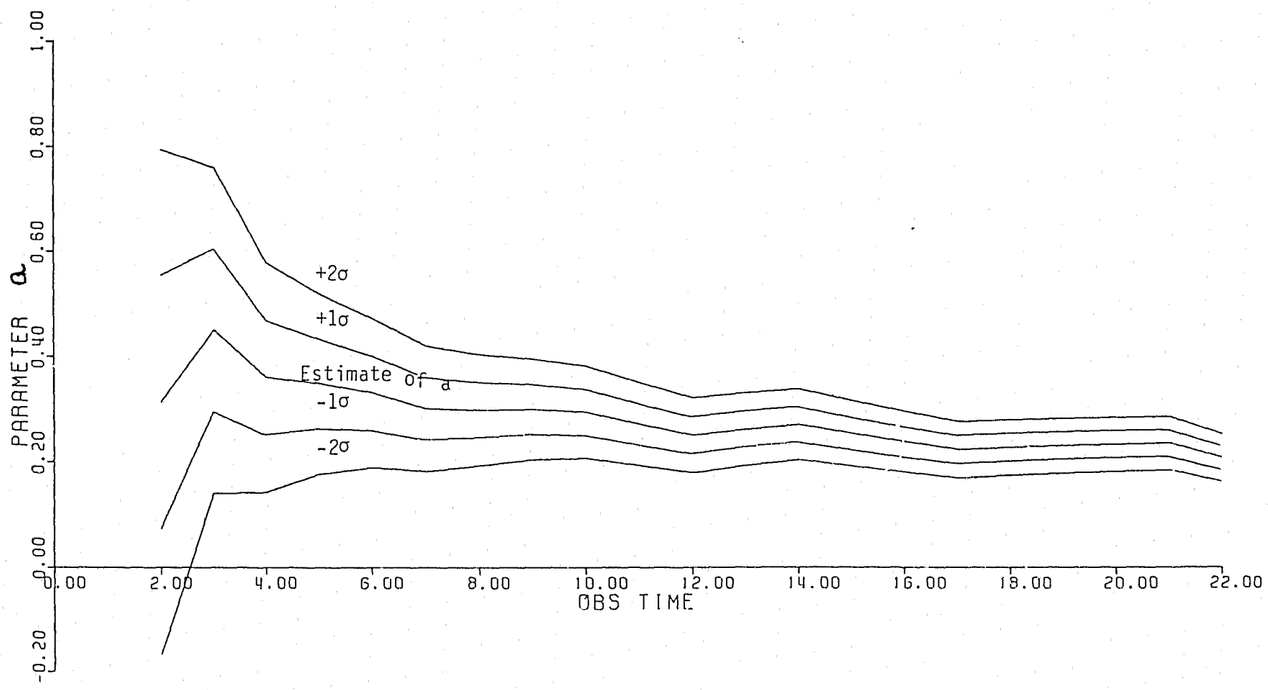


Figure 9-4 Variations in the Estimate of a as the Observation Time Increases

Comparing Models: North Carolina Data

Another aspect of model validation was brought out in our analysis of data from the North Carolina Department of Corrections, previously studied by Witte and Schmidt (1977). They used a truncated lognormal model of recidivism to fit the data.

We compared the lognormal model to the split population and mixed exponential models in the analysis of this data set.[5] Figure 9-5 depicts the North Carolina data and the curves generated by the models (using the parameters that maximized the likelihood functions). The parameters and chi-squared statistics for both models are given in Table 9-2. As can be seen, both the split-population and mixed exponential models provide better fits than does the lognormal model.

It is interesting to compare the parameters for the split-population and mixed exponential models, for both the Illinois data (Table 9-1) and the North Carolina data (Table 9-2). Recall that both models assume that all individuals fail according to one of two (constant) failure rates, but for the split population model one of the failure rates is zero.

For the Illinois data the parameters for the two models are quite different, whereas they are strikingly similar for the North Carolina data. That is, for the Illinois data the fraction subject to rapid failure is 42 percent for the split-population model, but 54 percent for the mixed exponential model; and the rate at which this fraction fails is 21 percent per month and 30 percent per month, respectively.

CONTINUED

2 OF 4

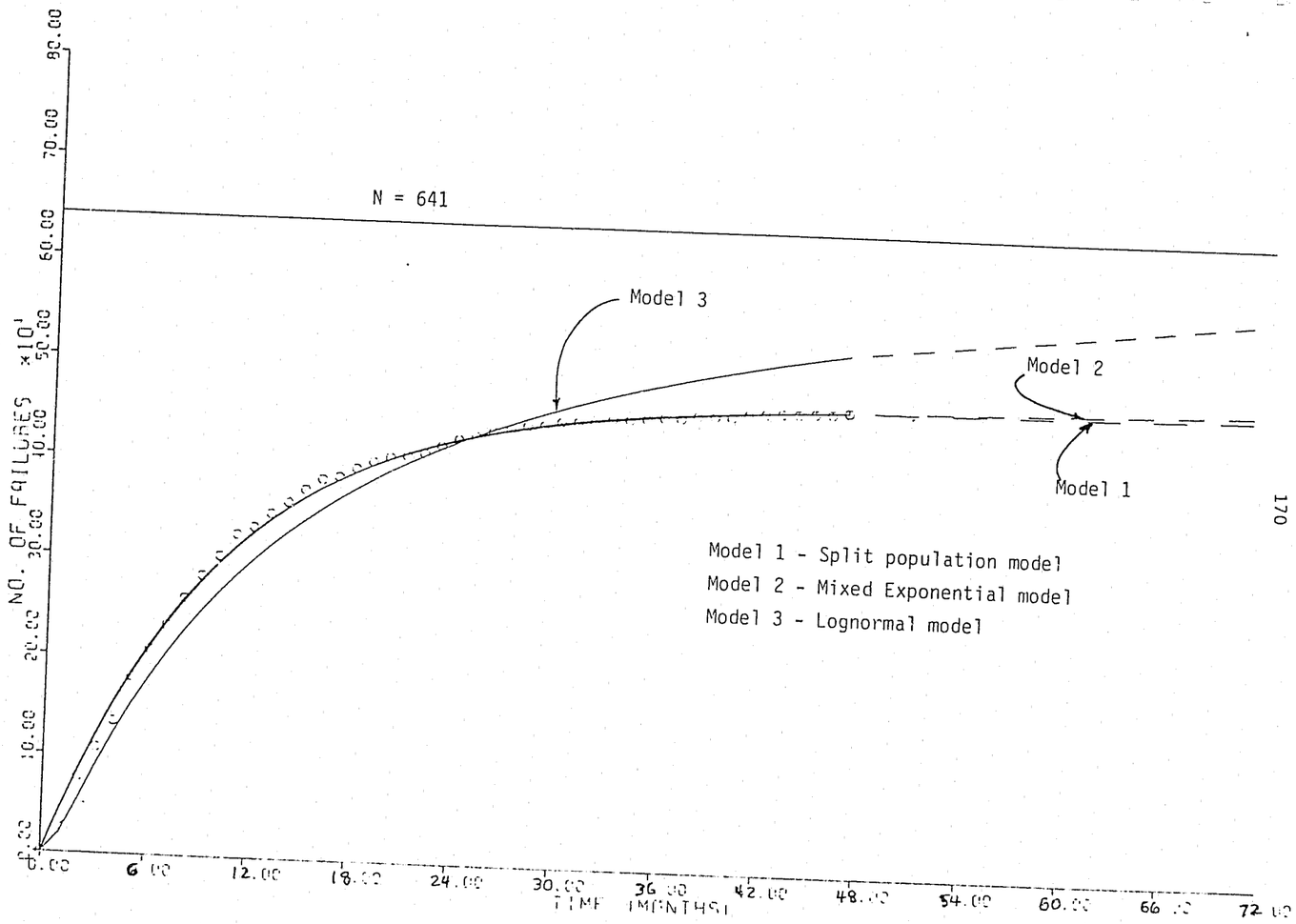


Figure 9-5 Fitting Models to North Carolina Data

Table 9-2 Three Models of Recidivism Compared to Data for 649 North Carolina Prison Releasees

Model 1. Split population exponential distribution: $F(t) = \gamma(1-\exp(-\phi t))$

Model 2. Mixed exponential distribution: $F(t)=\gamma(1-\exp(-\phi_1 t))+(1-\gamma)(1-\exp(-\phi_2 t))$

Model 3. Lognormal distribution: $F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^t \exp(-\frac{1}{2}(\frac{\ln y - \mu}{\sigma})^2)\, dy/y$

Model	Parameter estimates
1	$\gamma = 0.7329 \quad \phi = 0.09773$
2	$\gamma = 0.7127 \quad \phi_1 = 0.1022 \quad \phi_2 = 0.00142$
3	$\mu = 2.6068 \quad \sigma^2 = 1.9293$

Time interval (months)	Actual number failing during that interval		Forecast number of failures		
			Model 1	Model 2	Model 3
0 - 1	32		43.737	44.630	19.230
1 - 2	45		39.665	40.320	34.614
2 - 3	30		35.972	36.429	35.255
3 - 4	27		32.623	32.916	32.691
4 - 5	45		29.586	29.744	29.486
5 - 6	28		26.830	26.880	27.563
6 - 7	24		24.333	24.294	24.358
7 - 8	27		22.067	21.959	22.435
8 - 9	23		20. 13	19.851	20.512
9 - 10	20		18.149	17.948	18.589
10 - 11	22		16.459	16.229	17.307
11 - 12	8		14.927	14.678	16.025
12 - 13	14	} 26	25.814	25.289	28.204
13 - 14	12				
14 - 15	13	} 24	21.231	20.709	24.358
15 - 16	11				
16 - 17	3	} 12	17.461	16.975	21.794
17 - 18	9				
18 - 19	7	} 12	14.361	13.930	18.589
19 - 20	5				
20 - 21	2	} 5	11.812	11.448	16.666
21 - 22	3				
22 - 23	8	} 15	9.714	9.425	15.384
23 - 24	7				
24 - 25	4	} 9	7.990	7.775	12.820
25 - 26	5				
26 - 27	3	} 7	6. 71	6.430	12.179
27 - 28	4				
28 - 29	1	} 5	5.405	5.333	10.897
29 - 30	4				
30 - 31	1	} 2	4.445	4.439	10.256
31 - 32	1				
32 - 33	3	} 3	3.656	3.710	8.974
33 - 34	0				
34 - 35	1	} 3	3.007	3.115	8.333
35 - 36	2				
36 - 37	0				
37 - 38	3	} 3	4.507	4.864	14.743
38 - 39	0				
39 - 40	0				
40 - 41	3				
41 - 42	1				
42 - 43	1	} 5	3.049	3.559	12.820
43 - 44	0				
44 - 45	1				
45 - 46	0	} 3	1.620	2.081	10.897
46 - 47	2				
Over 47	176		175.997	176.040	116.021
Value of chi-square statistic			34.842	31.441	114.686
Degrees of freedom			26	25	26

We can contrast this with the parameter estimates for the North Carolina data. The fraction subject to rapid failure is 73 percent and 71 percent, respectively, for the two models; and the rate at which this fraction fails is 9.7 percent per month and 10.2 percent per month, respectively. Thus, for the North Carolina data there is hardly any difference between the two models; in fact, projecting these two models out to 72 months results in only a one percent difference between them.

We see, then, that the split population model cannot be applied unthinkingly to all data sets.[6] However, one can learn a good deal about the model's applicability by using it to study a variety of cohorts. The next two sections describe its application to cohorts provided by the U.S. Parole Commission and the U.S. Bureau of Prisons.

Forecasting Failure: U.S. Parole Commission Data

Shortly after we began our research on recidivism we met with the Research Unit of the U.S. Parole Commission. They had just completed a six-year follow-up study of 1806 federal parolees released in 1970 (Hoffman & Stone-Meierhoefer, 1977). Their data allowed us to put our methods to a rigorous test of their extrapolation ability. The Research Unit sent us data, but not all at once. Instead, we used the following arm's-length procedure.

We were sent data on failures[7] during the first six months of follow-up for four cohorts: "very good risks", "good risks", "fair risks", and "poor risks". An individual's risk level was estimated using the "Salient Factor Score" (Hoffman and Beck, 1974), a score based on the individual's prior criminality, education, employment and personal characteristics. Scores can range from zero to eleven (Figure 9-6). Very good risks were those with salient factor scores between 9 and 11; good risks, 6-8; fair risks, 4-5; and poor risks, 0-3. Based on the first six months of data we estimated the two parameters of the split population model for each risk group, and calculated the expected cumulative number of failures every six months over the remaining five and one-half years.[8]

This arm's-length procedure of data analysis was repeated twice. That is, failure data for months 7-12 were made available to us, based on which we made new estimates; and again for months 13-18. Following this we were given the data for the entire six years so that we could compare our predictions with the actual number of failures.

The data were not entirely accurate because of the computation method used to calculate the time to failure. The number of months to failure was calculated by subtracting the month of release from the failure month. This procedure does not produce the correct number of months at all times. Consider a person who is released on January 31 and fails one day later: he would be recorded as being out one month before failing. Then consider a person who is released on January 1

Case Name _____ *Register Number* _____

Item A .....

No prior convictions (adult or juvenile) = 2
One or two prior convictions = 1
Three or more prior convictions = 0

Item B .....

No prior incarcerations (adult or juvenile) = 2
One or two prior incarcerations = 1
Three or more prior incarcerations = 0

Item C .....

Age at first commitment (adult or juvenile): 18 years or older = 1
Otherwise = 0

Item D .....

Commitment offense did not involve auto theft = 1
Otherwise = 0

Item E .....

Never had parole revoked or been committed for a new offense while on parole = 1
Otherwise = 0

Item F .....

No history of heroin, cocaine, or barbiturate dependence = 1
Otherwise = 0

Item G .....

Has completed 12th grade or received GED = 1
Otherwise = 0

Item H .....

Verified employment (or full-time school attendance) for a total of at least 6 months during the last 2 years in the community = 1
Otherwise = 0

Item I .....

Release plan to live with spouse and/or children = 1
Otherwise =0

TOTAL SCORE .....

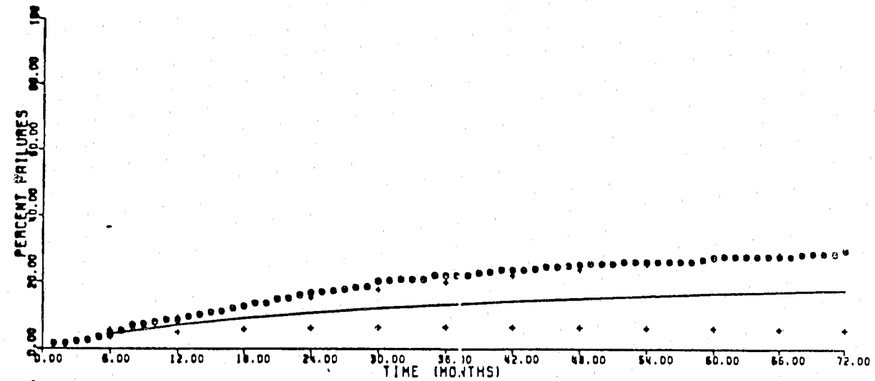
Figure 9-6 Salient Factor Score Sheet

and fails thirty days later: he would be recorded as being out zero months before failing. Therefore, certain inaccuracies were present.

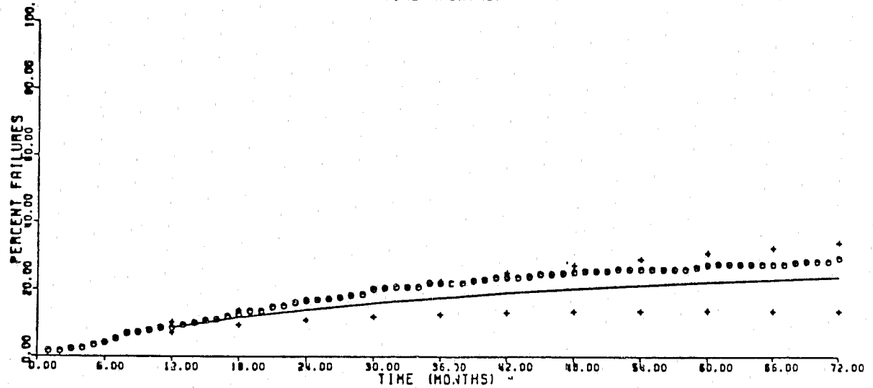
Figures 9-7 to 9-10 show the forecasts we provided.[8] The top curve in each of these figures shows the estimates out to six years using six months of data; the middle curve, 12 months; the bottom curve, 18 months. The actual data are represented by circles; the estimate is represented by a solid line; and the one-standard-deviation bounds on the estimate are represented by + symbols. As can be seen, the actual experience of the very good, good, and fair risk cohorts are all within one standard deviation of the estimates at 72 months. The estimator does not do as well for the poor-risk cohort; the twelve- and eighteen-month estimates fall well below the actual data.

To some extent, the relatively poor fit in the case of the "poor-risk" cohort points out the limits of applicability of the split-population model. This result might have been anticipated. Recall that in the split-population model we assume that one segment of the population will be rehabilitated, that this segment will revert to an "ambient" failure rate of zero. However, it appears that the poor-risk cohort is comprised of such poor risks that their ambient failure rate cannot be assumed to be zero, as is the case with the better-risk cohorts.

(a)



b)



(c)

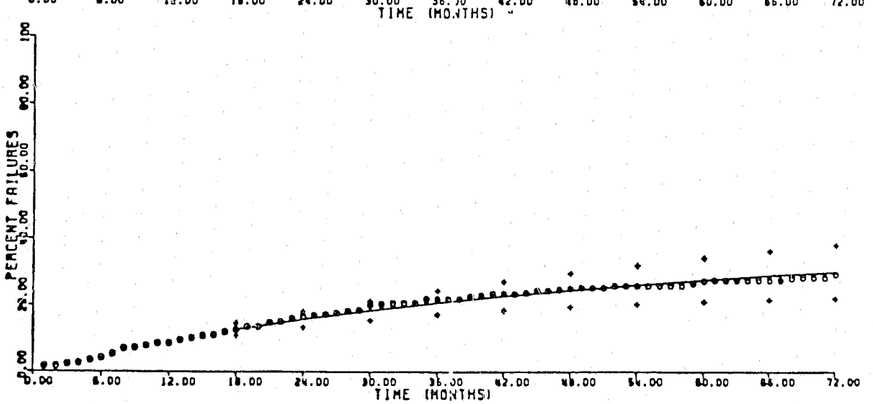
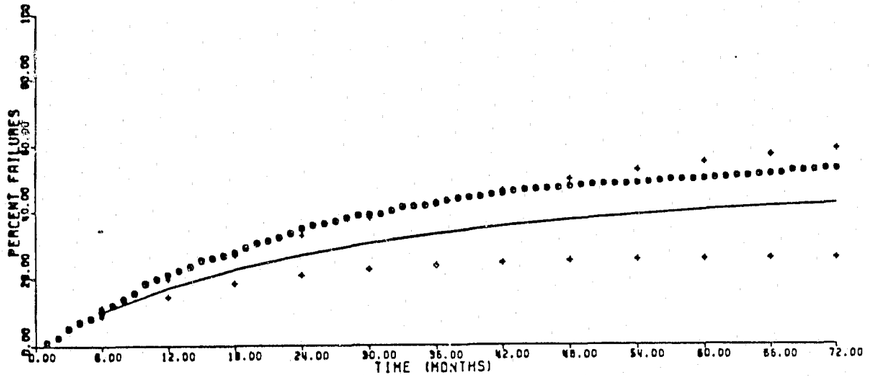


Figure 9-7 Failure Estimation for a "Very Good Risk" Cohort (N=319)

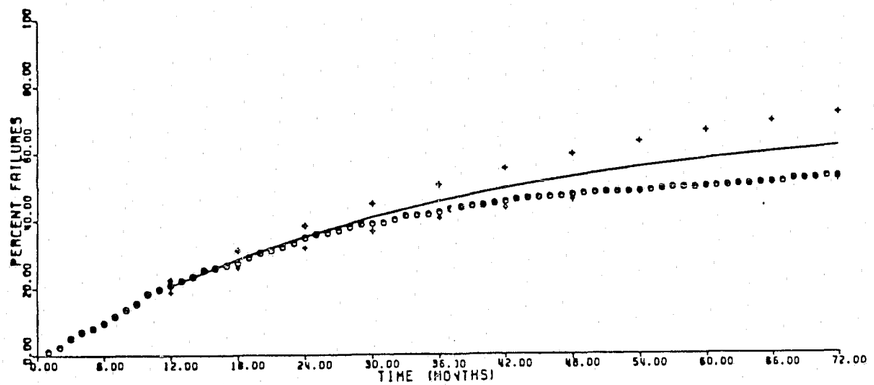
(a) 6-month estimate; (b) 12-month estimate; (c) 12-month estimate

NOTE: "—" represents the estimate, "+" one standard deviation

(a)



(b)



(c)

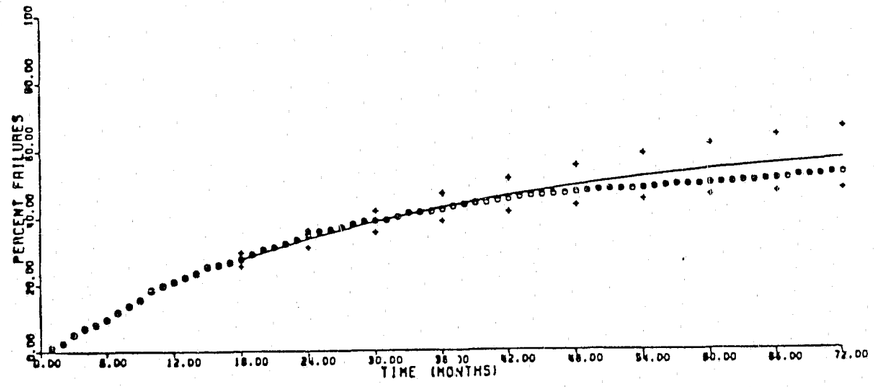
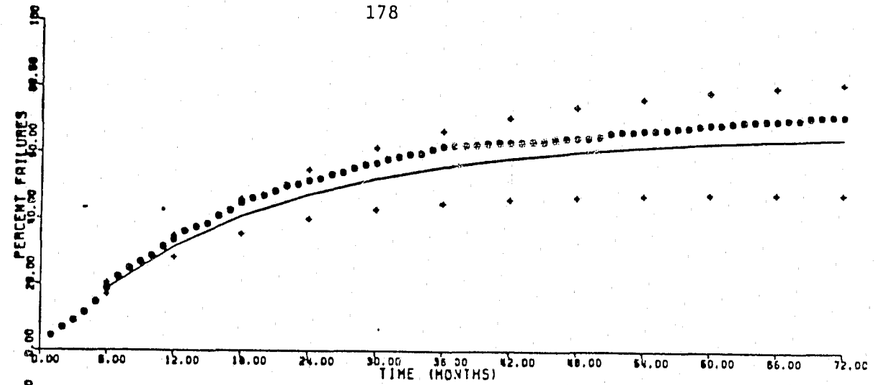


Figure 9-8 Failure Estimation for a "Good Risk" Cohort (N=483)

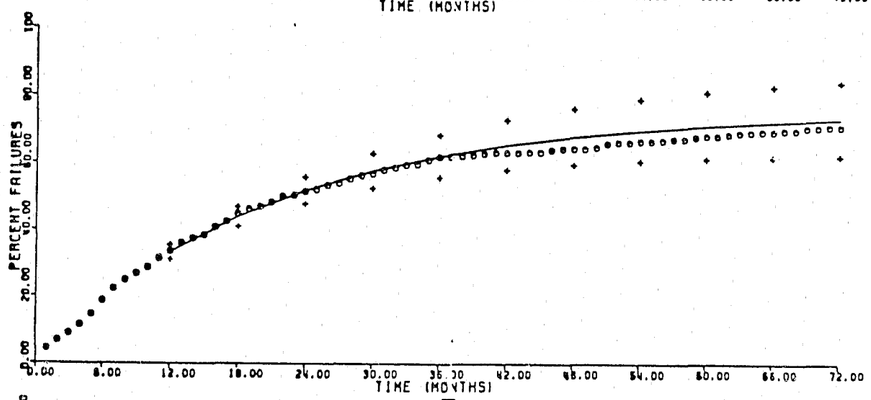
(a) 6-month estimate; (b) 12-month estimate; (c) 18-month estimate

NOTE: "—" represents the estimate, "+" one standard deviation

(a)



(b)



(c)

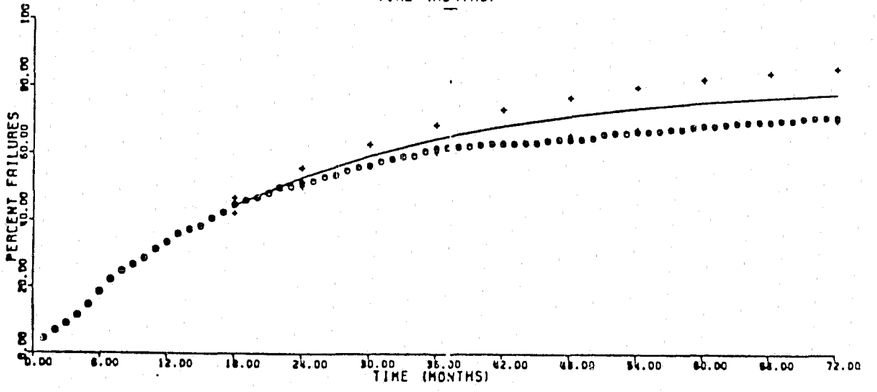


Figure 9-9 Failure Estimation for a "Fair Risk" Cohort (N=472)

(a) 6-month estimate; (b) 12-month estimate; (c) 18-month estimate

NOTE: "—" represents the estimate, "+" one standard deviation

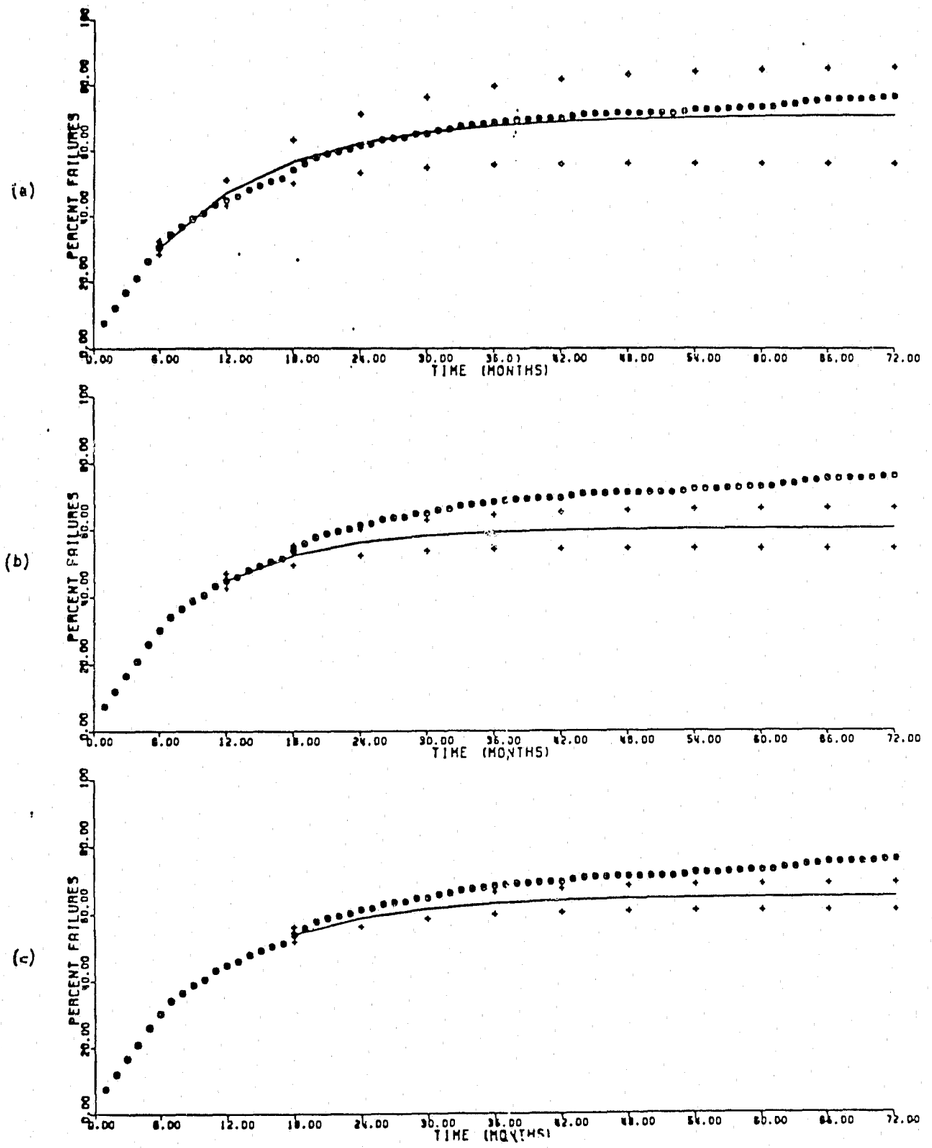


Figure 9-10 Failure Estimation for a "Poor Risk" Cohort (N=532) .

(a) 6-month estimate; (b) 12-month estimate; (c) 18-month estimate

NOTE: "—" represents the estimate, "+" one standard deviation

Towards a Covariance Model

In 1977 the U.S. Bureau of Prisons published the results of an eighteen-year follow-up study of federal prisoners released in 1956 (Kitchener, Schmidt and Glaser, 1977). The original sample selected in 1958 was of 1015 subjects, a ten percent sample of the 1956 male releasees. The Research Division of the Bureau of Prisons was able to obtain follow-up data on over ninety percent of this sample through the end of 1974. The cumulative number of failures was calculated for this cohort over the entire follow-up period, and turned out to be very consistent with the split-population model (Figure 7-5).

A failure was operationally defined as someone returned to prison as a parole violator or receiving any new sentence for a felony or felony-like offense (R). Other operational definitions were also
V
employed, but seemed to make little difference (Kitchener et al, 1977: 11).

The data included many variables describing each individual's criminal, educational, employment and personal history. From this information we were able to estimate the salient factor score (Figure 9-6) for each individual. We then considered all individuals with the same salient factor score to be in separate subgroups -- twelve in all -- and estimated the two parameters of the split-population model for each subgroup, as well as their covariances.

The results are shown in Figure 9-11, with ellipses representing the one standard deviation contours.[9] A general trend is discernible from this figure, represented by the dotted line: as the salient factor score decreases (and the "risk level" increases), there is a linear relationship between the fraction of failures and the failure rate, to a point. Beyond this point the fraction of failures does not change greatly while the failure rate continues to increase.

This relationship between the "risk level" and the parameters of the split-population model is worth noting for a number of reasons. The fact that the relationship is nonlinear means that ordinary linear regression techniques should not be used. It also suggests that the salient factor score is a reasonable predictor of recidivism parameters, one that should be considered further.

However, we do not feel it worthwhile continuing the investigation using this data set. It is atypical in that the prisoners were released from prison over twenty years ago; and the fact that they were released from federal prisons means that they are considerably different from state prisoners. For example, there are likely to be significant numbers of moonshiners and white-collar criminals within the group, as well as those who committed other crimes found only in the federal criminal code. This would limit the applicability of findings about this data set to state correctional systems.

Letter	Salient Factor Score	Risk Level
A	0*	High
B	1	.
C	2	.
D	3	.
E	4	.
F	5	.
G	6	.
H	7	.
I	8	.
J	9	.
K	10	.
L	11	Low

* This category omitted; too few in sample.

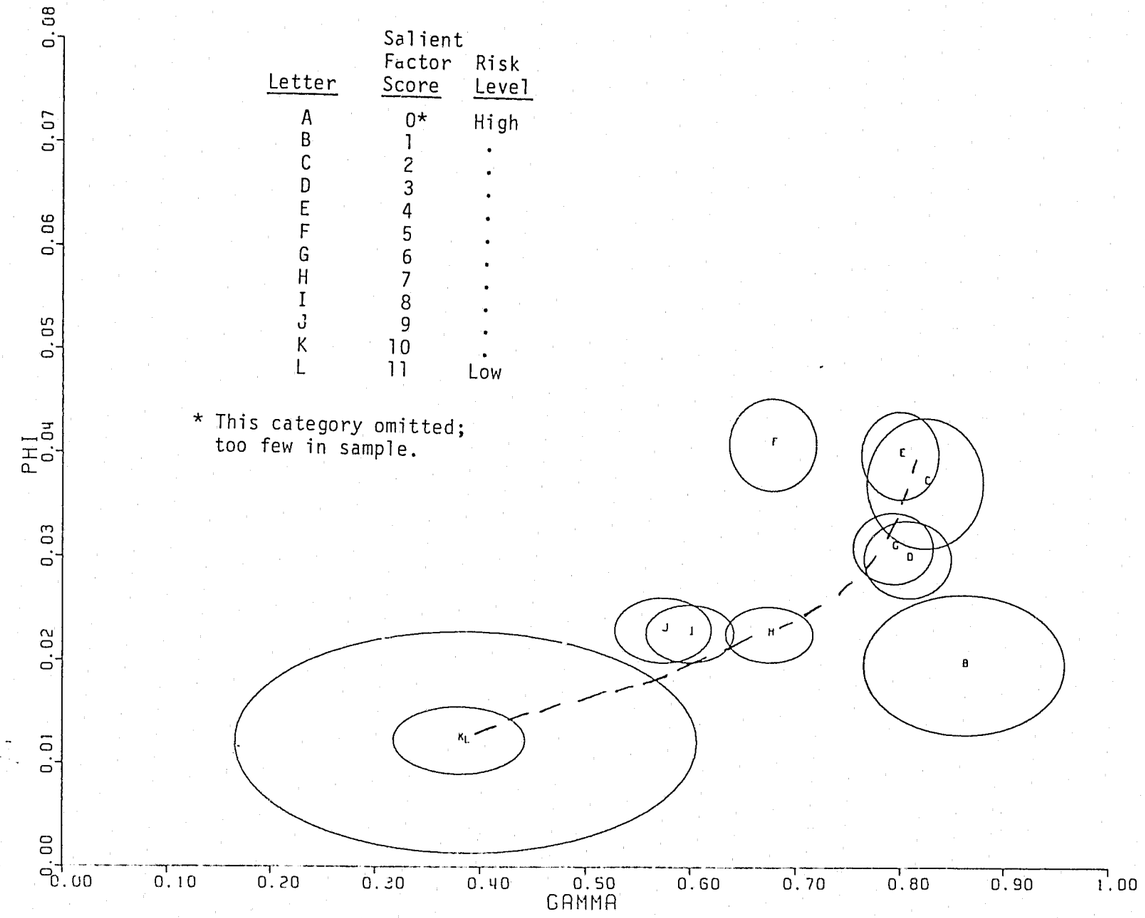


Figure 9-11 Estimates of γ and ϕ (and their 1σ contours for the US Bureau of Prisons Data, by Salient Factor Score

State-to-State Variation: Uniform Parole Reports

As an additional application of these techniques, data on parolees were obtained from the Uniform Parole Reporting Program, for thirty states, Puerto Rico and the District of Columbia. For each state the data contained the first change in disposition (e.g., discharge, arrest) for each individual in a cohort of people placed on parole in 1976. Each state used a follow-up time of one year for these parolees. Not all of the parolees were followed up for the full year: some were discharged before the end of a year, so track was lost of them; and others died before the full year was up. Thus, these individuals had follow-up times of less than one year.

Figure 9-12 shows a typical data sheet for a state. Note that there are seven different disposition categories:

- o Discharged
- o Returned to prison for a technical violation -- no new convictions and not in lieu of prosecution
- o Returned to prison for a technical violation -- new minor or lesser conviction(s) or in lieu of prosecution
- o Returned to prison for a technical violation -- in lieu of prosecution for new major offense(s)
- o Returned to prison -- no violation
- o Recommitted to prison with new major conviction(s)
- o Death

ROW PCT / COL PCT	DISCHARGED (1)	RET-NO CON;NO LU (2)	RET-MIN; IN LIEU (3)	RET-IN LIEU MAJ (4)	RET-NO VIOLATION (5)	RECOMMIT -NEW MAJ (6)	DEATH (7)	ROW TOTAL
MONTHS								
0	7	0	0	0	0	1	0	8
LESS THAN ONE	87.5	0.0	0.0	0.0	0.0	12.5	0.0	4.0
	6.0	0.0	0.0	0.0	0.0	2.1	0.0	
1	3	5	1	1	0	2	1	13
ONE	23.1	38.5	7.7	7.7	0.0	15.4	7.7	6.4
	2.6	17.9	20.0	100.0	0.0	4.2	33.3	
2	4	5	0	0	0	2	0	11
TWO	36.4	45.5	0.0	0.0	0.0	18.2	0.0	5.4
	3.4	17.9	0.0	0.0	0.0	4.2	0.0	
3	2	3	1	0	0	2	0	8
THREE	25.0	37.5	12.5	0.0	0.0	25.0	0.0	4.0
	1.7	10.7	20.0	0.0	0.0	4.2	0.0	
4	2	6	0	0	0	4	0	12
FOUR	16.7	50.0	0.0	0.0	0.0	33.3	0.0	5.9
	1.7	21.4	0.0	0.0	0.0	8.3	0.0	
5	6	1	1	0	0	6	0	14
FIVE	42.9	7.1	7.1	0.0	0.0	42.9	0.0	6.9
	5.2	3.6	20.0	0.0	0.0	12.5	0.0	
6	7	2	0	0	1	7	0	17
SIX	41.2	11.8	0.0	0.0	5.9	41.2	0.0	8.4
	6.0	7.1	0.0	0.0	100.0	14.6	0.0	
7	17	1	1	0	0	4	0	23
SEVEN	73.9	4.3	4.3	0.0	0.0	17.4	0.0	11.4
	14.7	3.6	20.0	0.0	0.0	8.3	0.0	
8	19	1	0	0	0	4	1	25
EIGHT	76.0	4.0	0.0	0.0	0.0	16.0	4.0	12.4
	16.4	3.6	0.0	0.0	0.0	8.3	33.3	
9	10	1	1	0	0	7	0	19
NINE	52.6	5.3	5.3	0.0	0.0	36.8	0.0	9.4
	8.6	3.6	20.0	0.0	0.0	14.6	0.0	
10	11	1	0	0	0	3	0	15
TEN	73.3	6.7	0.0	0.0	0.0	20.0	0.0	7.4
	9.5	3.6	0.0	0.0	0.0	6.3	0.0	
11	12	2	0	0	0	2	0	16
ELEVEN	75.0	12.5	0.0	0.0	0.0	12.5	0.0	7.9
	10.3	7.1	0.0	0.0	0.0	4.2	0.0	
12	16	0	0	0	0	4	1	21
TWELVE	76.2	0.0	0.0	0.0	0.0	19.0	4.8	10.4
	13.8	0.0	0.0	0.0	0.0	8.3	33.3	
COLUMN TOTAL	116	28	5	1	1	48	3	202
	57.4	13.9	2.5	0.5	0.5	23.8	1.5	100.0

RAW CHI SQUARE = 106.70062 WITH 72 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0040
CRAMER'S V = 0.29671
CONTINGENCY COEFFICIENT = 0.58792
LAMBDA (ASYMMETRIC) = 0.06215 WITH MONTHS DEPENDENT. = 0.09302 WITH REMOV DEPENDENT.
LAMBDA (SYMMETRIC) = 0.07224
UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.08729 WITH MONTHS DEPENDENT. = 0.19206 WITH REMO
UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.12003
KENDALL'S TAU B = -0.16643. SIGNIFICANCE = 0.0017
KENDALL'S TAU C = -0.14307. SIGNIFICANCE = 0.0017
GAMMA = -0.22220
SOMERS'S D (ASYMMETRIC) = -0.20657 WITH MONTHS DEPENDENT. = -0.13408 WITH REMOV DEPENDEN
SOMERS'S D (SYMMETRIC) = -0.16262
ETA = 0.35713 WITH MONTHS DEPENDENT. = 0.27215 WITH REMOV DEPENDENT.
PEARSON'S R =-0.09576 SIGNIFICANCE = 0.0876

NUMBER OF MISSING OBSERVATIONS = 339

Figure 9-12 UPR Data for a State

For the purposes of our analysis we used R_V as the operational definition of recidivism. The categories "discharge" and "death" were lumped together into one category (those who had not failed when track was lost of them), and the remaining five dispositions formed the failure category. For the cohort of Figure 9-12, then, we see that one parolee failed in the same month that he was released, nine failed in the first month after release, etc. In addition, seven parolees were observed for less than one month (without failing) before track was lost of them, four parolees were dropped from the cohort (without having failed) in the first month following their parole, etc.

New dispositions were recorded for 541 persons. In addition, 339 persons in the cohort had no change in their dispositions when the one-year follow-up period was completed. We can therefore consider them as having been lost track of after twelve months, and add them to the seventeen who dropped out of the cohort in the twelfth month. Table 9-3 lists the data for every state.

Likelihood functions were generated for each of these cohorts. They are depicted in Appendix B. As can be seen, in a number of cases there is no local maximum -- Delaware, Ohio, Puerto Rico, South Carolina, South Dakota, Wyoming and Utah -- in others there is a fairly well-defined maximum -- Florida, Illinois, Kansas, Kentucky, Missouri, Nebraska, New Jersey, New Mexico, New York, Pennsylvania, Texas, Wisconsin and Virginia -- and in the remaining states there is a maximum but it is not very well-defined.

State		Month 1	2	3	4	5	6	7	8	9	10	11	12
Alabama	F	5	4	7	5	7	7	3	5	8	0	1	1
	X	1	4	5	10	11	16	14	20	12	14	12	175
Alaska	F	1	0	1	1	1	1	1	0	0	0	0	0
	X	0	0	1	2	1	0	0	2	0	0	0	14
Delaware	F	0	1	1	0	2	0	1	2	1	0	1	0
	X	0	0	3	3	1	2	5	3	0	3	3	53
D.C.	F	2	5	2	6	6	4	10	6	4	4	1	1
	X	1	6	4	1	0	4	5	2	4	2	3	254
Florida	F	36	26	37	38	35	24	20	29	16	4	3	2
	X	5	4	16	19	54	59	56	59	68	62	57	2160
Georgia	F	3	7	4	8	7	6	8	3	2	1	3	1
	X	2	13	24	47	46	29	23	16	20	17	14	208
Idaho	F	2	5	0	2	3	2	3	5	0	1	0	0
	X	1	0	1	1	0	1	2	4	2	2	4	144
Illinois	F	61	40	42	35	40	24	20	19	10	16	17	7
	X	2	3	0	3	1	2	2	1	2	6	2	1838
Iowa	F	5	7	6	8	5	6	5	2	5	3	2	2
	X	0	1	0	0	1	2	1	0	2	2	5	193
Kansas	F	47	28	24	15	19	11	10	10	8	6	5	2
	X	0	0	0	1	1	1	0	0	0	2	0	634
Kentucky	F	39	33	21	22	26	22	10	12	7	7	4	6
	X	1	2	6	16	38	48	49	18	6	17	19	863
Maine	F	10	7	6	10	8	10	6	5	9	4	4	4
	X	11	4	2	2	6	7	17	20	10	11	12	356
Michigan	F	20	11	14	12	12	11	10	10	7	11	11	2
	X	2	1	2	6	9	17	14	15	16	10	11	395
Missouri	F	16	8	8	1	5	5	3	0	2	1	1	1
	X	0	0	0	2	0	2	3	5	14	20	28	182
Nebraska	F	7	5	9	5	5	5	4	2	1	0	0	0
	X	12	12	13	13	19	8	6	10	12	13	20	130
New Hamp.	F	1	1	1	3	1	2	1	1	1	1	2	0
	X	0	0	1	1	0	0	2	0	3	1	1	4

State		Month 1	2	3	4	5	6	7	8	9	10	11	12
New Jersey	F	42	38	35	23	23	16	10	3	5	1	2	2
	X	10	14	13	30	17	20	13	22	25	17	12	1140
New Mexico	F	18	8	12	8	9	4	8	4	5	6	4	6
	X	1	2	1	2	3	1	4	1	1	1	1	245
New York	F	9	2	6	9	4	2	2	4	3	1	2	2
	X	1	0	1	10	0	0	1	1	1	0	1	215
N. Dakota	F	0	0	1	2	0	0	0	0	0	0	0	0
	X	5	6	7	5	9	3	6	1	6	3	4	52
Ohio	F	4	3	3	7	6	7	7	5	9	7	2	0
	X	0	1	0	1	0	0	1	0	0	1	0	364
Penna.	F	13	23	15	13	14	12	12	7	16	11	10	5
	X	14	10	10	16	14	17	18	15	22	19	31	2420
Pto. Rico	F	0	0	0	0	3	1	2	1	2	2	1	3
	X	0	0	1	0	1	0	0	0	6	1	3	311
S. Car.	F	2	5	2	11	12	12	11	7	9	6	5	0
	X	2	6	8	11	15	31	24	23	24	36	19	625
S. Dakota	F	2	2	1	1	1	0	0	4	0	1	0	1
	X	2	2	8	18	11	7	12	9	5	7	6	59
Tennessee	F	5	3	7	4	7	5	2	3	2	3	5	2
	X	0	0	0	1	2	1	2	4	1	1	1	1239
Texas	F	63	57	69	69	61	41	43	37	37	26	25	14
	X	29	72	83	99	115	108	123	142	162	186	190	2472
Utah	F	0	5	1	3	8	5	3	2	3	6	7	1
	X	3	0	0	0	1	0	2	0	1	1	0	142
Virginia	F	38	22	34	23	21	27	19	16	12	10	6	2
	X	4	0	2	0	4	3	2	4	2	2	44	1496
W. Va.	F	2	3	0	5	2	2	2	5	3	3	0	1
	X	1	1	0	0	1	0	1	1	1	0	1	314
Wisconsin	F	7	9	5	7	5	8	3	6	4	7	1	0
	X	2	5	4	5	11	14	2	13	9	8	5	389
Wyoming	F	0	0	0	0	4	5	1	2	1	1	2	1
	X	1	0	0	0	0	0	0	0	0	0	0	27

"F": Failure data (Number failing in each month)

"X": Exposure data (Number lost track of in each month)

Table 9-3 State Data Reported to the UPR Program

The variations we note here in no way reflect on the quality of the data or on the nature of the parole process in the states. They may be caused by differences in the parole or correctional agencies' policies or structure; they may reflect differences in parolee characteristics; they may reflect differences in the way data were reported to the UPR Program. The lack of a well-defined maximum may mean that the sample size is too small; it may mean that a longer observation time is needed; or it may mean that the split-population model is inappropriate.

Figure 9-13 is a plot of the maximum likelihood estimates of γ (the fraction of the cohort expected to fail eventually) and q (the proportion of to-be failures who survive to the next month) for the likelihood functions having local maxima (i.e., excepting states without local maxima). The values are given in Table 9-4, along with lower bound estimates of the parameters' standard deviations. These "standard deviations" are useful in those cases where the maximum is well-defined, since the likelihood functions closely resemble bivariate normal distributions. In such cases normal statistical tests, such as the ones proposed by Lloyd and Joe (1979) are appropriate and accurate.

The estimates of the standard deviation become less and less appropriate and accurate as the maximum becomes less and and less well-defined, especially for γ -- the variation in the q direction

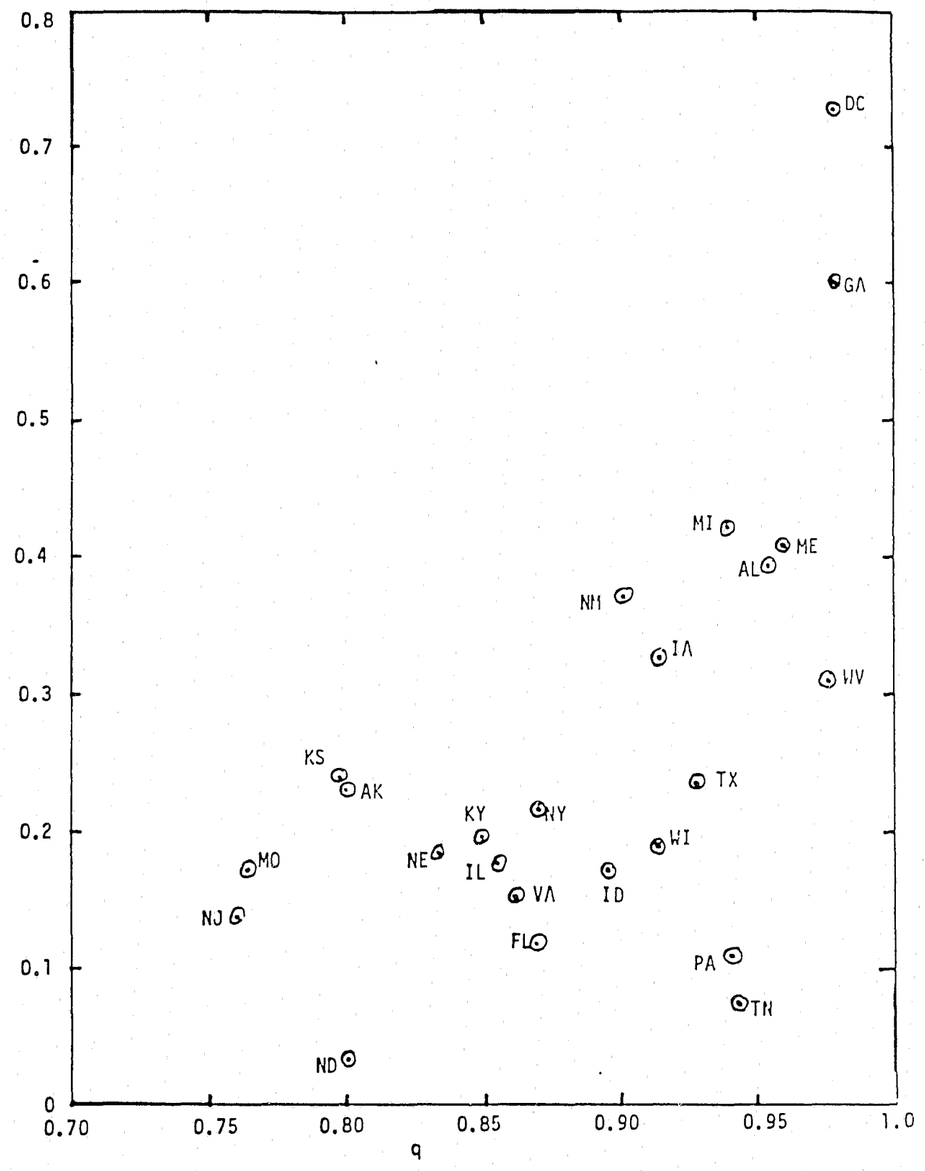


Figure 9-13 Maximum Likelihood Estimates of γ and q for States, Using UPR Data

Table 9-4 Maximum Likelihood Estimates of γ and q for States, Using UPR Data

State	γ	σ_γ	q	σ_q
AL	0.393	0.256	0.952	0.040
AK	0.233	0.102	0.799	0.129
DE	no local maximum			
DC	0.727	1.274	0.980	0.040
FL	0.120	0.010	0.870	0.017
GA	0.600	1.026	0.979	0.040
ID	0.174	0.060	0.896	0.057
IL	0.179	0.011	0.855	0.015
IA	0.326	0.089	0.914	0.037
KS	0.241	0.016	0.798	0.020
KY	0.198	0.016	0.849	0.019
ME	0.407	0.237	0.957	0.031
MI	0.421	0.126	0.940	0.025
MO	0.175	0.023	0.765	0.040
NE	0.183	0.032	0.834	0.044
NH	no local maximum			
NJ	0.140	0.009	0.760	0.020
NM	0.371	0.066	0.901	0.028
NY	0.218	0.044	0.880	0.040
ND	0.035	0.023	0.808	0.173
OH	no local maximum			
PA	0.108	0.030	0.941	0.023
PR	no local maximum			
SC	no local maximum			
SD	no local maximum			
TN	0.073	0.038	0.943	0.040
TX	0.236	0.028	0.928	0.012
UT	no local maximum			
VA	0.155	0.012	0.861	0.018
WV	0.310	0.581	0.975	0.054
WI	0.189	0.052	0.915	0.035
WY	no local maximum			

appears to more closely resemble a normal (or at worst a beta) distribution.

Summary

From the variety of applications that we have presented, it should be clear that no absolute statements can be made about specific models of the recidivism process. No one model is clearly superior for all cohorts, nor is there any simple test to use to select the best model.

One can, however, make certain generalizations about model selection and testing. First, the split population model seems to be the best one to use initially. Although there are models with better overall fits to the existing data, this model may be better for forecasting than it is for curve-fitting. In addition, it is a simple model for which maximum likelihood estimates of the parameters can often be found graphically, when the observation time is the same for all non-failures. Knowing the values of K, T, N, and n (see Chapter 8 for their definitions), one can obtain estimates of γ and q graphically.

Second, the mixed exponential is another useful model. But it may be best to apply this model only in cases where the empirical distribution of failures seems to be asymptotic to a sloped line (as in Figure 9-10) rather than a horizontal line. Otherwise the extra

parameter (the second failure rate) can be misleading. For example, I would expect the split population model to be a better predictor in Figure 9-2 than the mixed exponential model.

A suitable means of testing models does not yet exist. The chi-squared goodness of fit test, which is the one used most often, does not test for predictive ability of models but for closeness to the data.

CHAPTER 10

CONCLUSION AND FUTURE RESEARCH

This research has attempted to clarify the the concept of recidivism, to make its definition and its analysis more precise. The first half of the report addressed the definitional problem, and resulted in the development of a preliminary taxonomy of recidivism definitions so that studies using dissimilar definitions would not be considered comparable.

But resolving the definitional problem will not by itself resolve the issue of comparability of recidivism statistics across jurisdictions. Each state is sui generis: A state which places ten percent of its convicted offenders on probation cannot be compared directly with one that places thirty percent on probation. A state with overcrowded prisons, in which the parole board and parole officers are urged to keep people on the street if at all possible, cannot be compared directly with a state that is closing prisons. A state with determinate sentencing cannot be compared directly with a state with indeterminate sentencing. And variations in recidivism statistics may be even greater between urban and rural jurisdictions within a state than between states. In other words, one should be extremely careful before making inferences based on raw recidivism rates, notwithstanding the use of the same definition of recidivism and of appropriate analytic techniques.

This should not be taken as a recommendation for a total proscription against the use of recidivism as a measure of correctional effectiveness. Carefully designed experiments (and quasi-experiments) will still be useful in making limited inferences about programs. And the applications suggested at the end of Chapter 3 and described in Chapter 9 will shed a great deal of light on the nature of offender behavior and the criminal justice system's response to it. But blanket statements about programs or states, so avidly sought by some journalists and policy-makers, can rarely be gleaned from the data and will usually be misleading.

The methods of analysis described and applied in Chapters 7 to 9 are not yet complete. We have gained a great deal of experience in testing them using almost forty data sets, and have learned of some of the limitations of the methods and the models on which they are based. We intend to address these limitations in subsequent researh.

We plan to continue our research in the following four directions: developing criteria for model selection, developing confidence intervals for the model parameters, developing covariate models, and investigating other models of recidivism.

Model Selection

The discussions of model selection in Chapters 7 and 9 show the limitations of the standard chi-square goodness-of-fit test in selecting the appropriate process model. Two alternative selection criteria were suggested in Chapter 9, both based on withholding later data and attempting to forecast results from them. In Figures 9-3 and 9-4 it was shown how the parameter estimates did not change greatly even if the Illinois cohort had only a six-month follow-up time. This was suggested in our original paper (Maltz and McCleary, 1977) and elaborated on subsequently (Miley, 1978; Maltz and McCleary, 1978). Criteria for selection of models can be developed based on such a procedure. We intend to investigate this possibility by generating estimates of parameters for other models than the split population model for the data sets analyzed herein.

A more useful criterion for model selection could be based on the ability of each model to forecast the future number of failures, as shown in Fiqures 9-7 to 9-10. We intend to employ the data sets made available us in this exercise as well. Although the model with the best forecasting ability will probably be the one with the most consistent parameter estimates, our method of model selection may have advantages over the other. We intend to explore methods more fully, with the expectation of developing rigorous tests of forecast quality for use in model selection.

Significance Tests and Confidence Intervals

Moments of the likelihood function of Equation 8-9 (and its variants) cannot always be used to develop confidence statements about the parameter estimates, because the likelihood function is not always distributed (nearly) normally. It is not presently clear under what conditions moments can be used. We intend to focus our attention on this problem in the forthcoming year.

Bayesian techniques have been used to overcome this problem (Maltz and Pollock, 1981) but they introduce other problems. In particular, they require that one specify a prior state of knowledge about the parameters, which in many cases can only be guessed at. Since this prior information is incorporated into the Bayesian estimation procedure, the resultant parameter estimates and confidence intervals can be faulted for having been based on guesswork. However, it may be that the results are largely insensitive to variations in the prior state of knowledge. This is an area we intend to explore more fully.

Covariate Models

In true experiments, the unit of analysis is the group. Experimental and control groups are selected in such a way (matching, randomizing) as to be equivalent, so that the treatment is the only factor which can lead to different outcomes. But this is rarely the case in correctional program evaluations, so the two groups are more

often than not dissimilar. Covariate models attempt to compensate for such dissimilarities by using the individual (or group of individuals with common characteristics) as the unit of analysis. Relevant characteristics of the individual are incorporated in the model -- e.g., age at first arrest, number of prior convictions, etc. One way of incorporating these characteristics is to consider each individual i to have a probability γ_i of ultimately recidivating, with

$$\log(\gamma_i) = b_0 + \sum_j b_j x_{ij}$$

where x_{ij} is the value of the jth characteristic for the ith individual and the b_j are multiplicative constants (to be estimated). A similar relationship can be posited for other parameters.

We have not investigated covariate models specifically in the research reported on herein. However, in our analysis of the US Bureau of Prisons data (Fiqure 8-11) we showed how the parameters γ and ϕ varied with the salient factor score. This type of analysis will pave the way for the development of covariate models. Data sets from North Carolina and Georgia, and others we may obtain from other states, will be used to develop such models.

- - - - - - -

It must be emphasized again that these methodological advances improve the utility of recidivism as a measure of effectiveness only insofar as the measure is applicable. Improved statistical techniques are worthless if they are applied to inappropriate data or if their results are improperly interpreted. Sir Josiah Stamp's (1929: 258) caution about official statistics is quite appropriate:

> "The individual source of the statistics may easily be the weakest link. Harold Cox tells a story of his life as a young man in India. He quoted some statistics to a Judge, an Englishman, and a very good fellow. His friend said, "Cox, when you are a bit older, you will not quote Indian statistics with that assurance. The Government are very keen on amassing statistics -- they collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But what you must never forget is that every one of those figures comes in the first instance from the chowty dar [village watchman], who just puts down what he damn pleases."

In this report we have prepared many "wonderful diagrams". Whether they have any significance (other than esthetic) is in the hands of those who furnish the data.

APPENDIX A

THE "CRITICAL TIME" MODEL OF RELEASEE BEHAVIOR

In Chapter 7 we described the split population distribution as generated by what might be called a "good guy, bad guy" model of releasee behavior. That model posits that each individual in a cohort will behave according to one of two modes of behavior. In one mode the individual will never fail (or will never be observed to fail); $1-\gamma$ is the probability that he will behave in this way. In the other mode (in which he will fall with probability γ) he will fail in such a way that his time of failure is an exponentially distributed random variable. We also discussed a variant of this model, in which the failure rate for an individual in the first mode is not zero, but is still much lower than the failure rate for the second mode. This lower failure rate can be considered the "ambient" failure rate, the rate at which the general population might fail.

We now show that the split population and mixed exponential distributions can be generated by another model of releasee behavior, termed the "critical time" model. We will first describe it for the mixed exponential case; extension to the split population case is straightforward.

This model assumes that each individual in the cohort is liable to failure until a critical time θ; as before, the time of failure is an exponentially distributed random variable. Call the individual's failure rate λ_1.

After the critical time has been reached, the individual's failure rate drops to λ_2, which is zero for the split population case, positive and much smaller than λ_1 for the mixed exponential case.[1] In other words, if the individual can make it on the outside until time θ without getting arrested his failure rate will then drop to the ambient failure rate (or zero, for the split population case) and he can be considered a success.

The conditional probability distribution (conditional on θ) for this model is

$$P_\theta(t) = \begin{cases} 1 - \exp(-\lambda_1 t) & 0 < t \leq \theta < \infty \\ 1 - \exp[-\lambda_1 \theta - \lambda_2 (t-\theta)] & 0 \leq \theta \leq t < \infty \end{cases} \qquad \text{(A-1)}$$

Now let us assume that the critical time θ is itself a random variable whose probability distribution is $G(\theta)$ and whose probability density function is $g(\theta) = dG(\theta)/d\theta$. The unconditional probability distribution of failure times is then

$$P(t) = \int_0^t P_\theta(t)\, g(\theta)\, d\theta \qquad \text{(A-2)}$$

$$= \int_0^t \{1-\exp[-\lambda_1\theta-\lambda_2(t-\theta)]\}g(\theta)d\theta + \int_t^\infty [1-\exp(-\lambda_1 t)]g(\theta)d\theta$$

If θ is distributed exponentially (i.e., if $g(\theta) = \mu \exp(-\mu\theta)$), then the solution to Equation A-2 is

$$P(t) = \frac{\lambda_1-\lambda_2}{\mu+\lambda_1-\lambda_2}\{1-\exp[-(\mu+\lambda_1)t]\} + \frac{\mu}{\mu+\lambda_1-\lambda_2}\{1-\exp(-\lambda_2 t)\} \qquad \text{(A-3)}$$

This distribution is equivalent to the mixed exponential distribution

$$P(t) = \gamma[1 - \exp(-\phi_1 t)] + (1-\gamma)[1-\exp(-\phi_2 t)] \qquad \text{(A-4)}$$

The relationship between the two sets of parameters is fairly simple:

$$\gamma = \frac{\lambda_1-\lambda_2}{\mu+\lambda_1-\lambda_2} \qquad \mu = (1-\gamma)(\phi_1-\phi_2)$$

$$\phi_1 = \mu + \lambda_1 \qquad \lambda_1 = \gamma\phi_1 + (1-\gamma)\phi_2$$

$$\phi_2 = \lambda_2 \qquad \lambda_2 = \phi_2$$

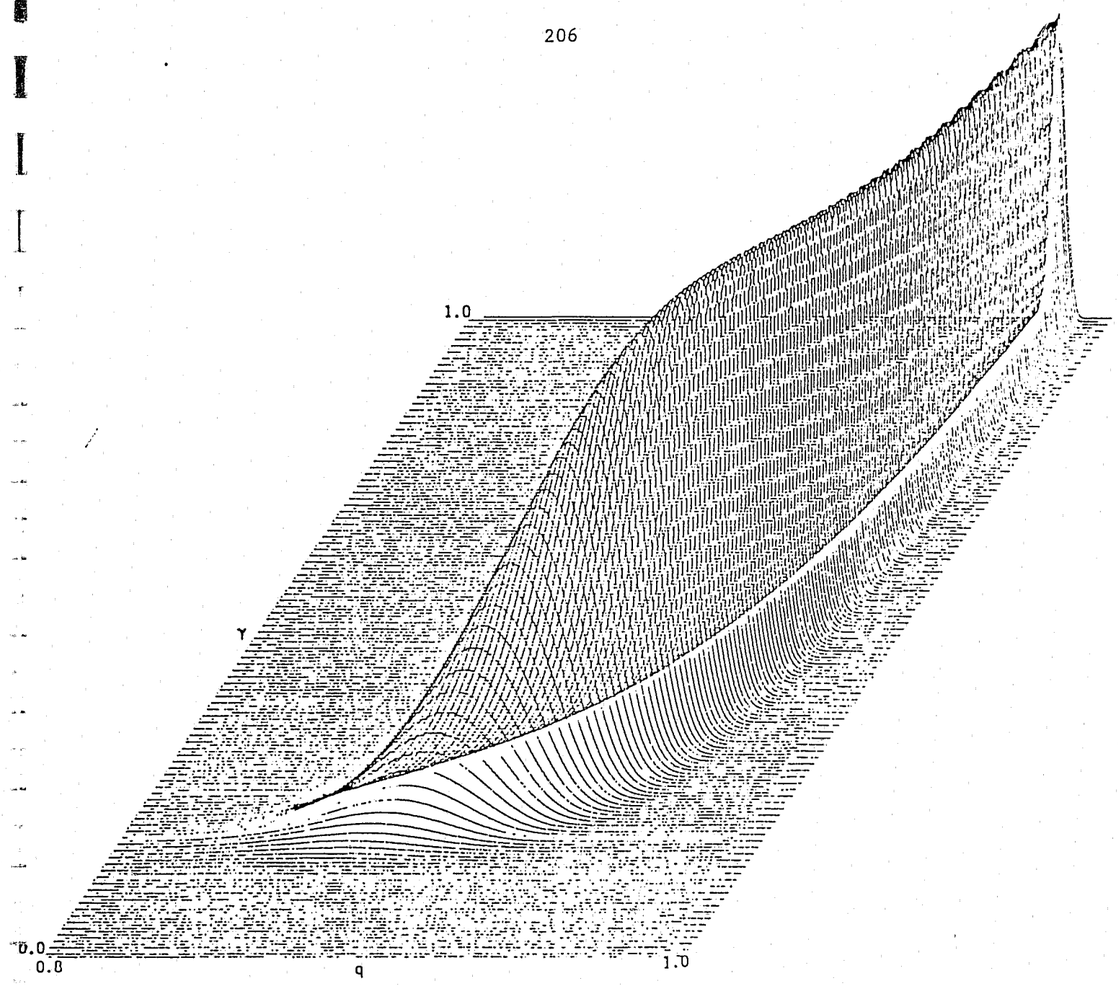
Setting λ_2 and ϕ_2 equal to zero changes Equations A-3 and A-4 into the split population distribution. Thus we see that when the generating processes are exponential, the "good guy, bad guy" model and the "critical time" model are exactly equivalent.

NOTE

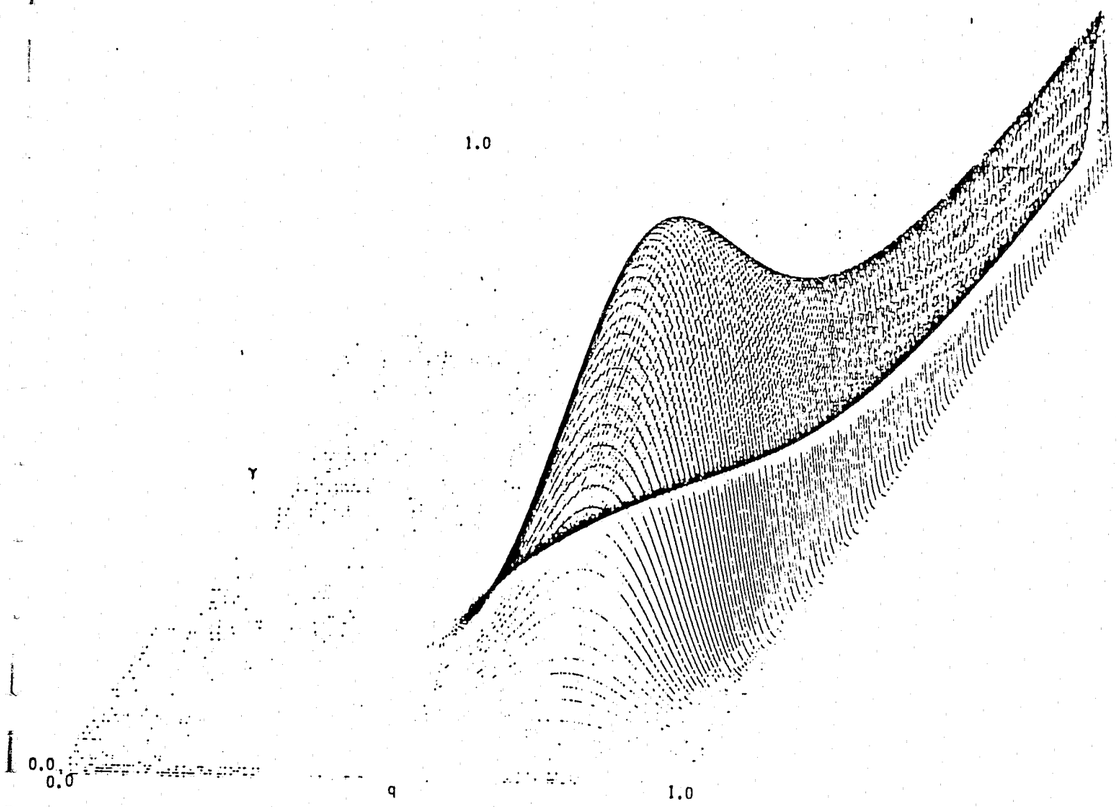
1. In Harris, Kaylan and Maltz (1981) we show that when a mixed distribution consists of two exponential distributions whose failure rates are of the same order of magnitude, one cannot easily distinguish it from a single exponential distribution whose failure rate is between the two failure rates. If the data are "noisy", distinguishing the two situations becomes even more diffiucult. Thus we state that the smaller failure rate must be much smaller than the larger failure rate.

APPENDIX B

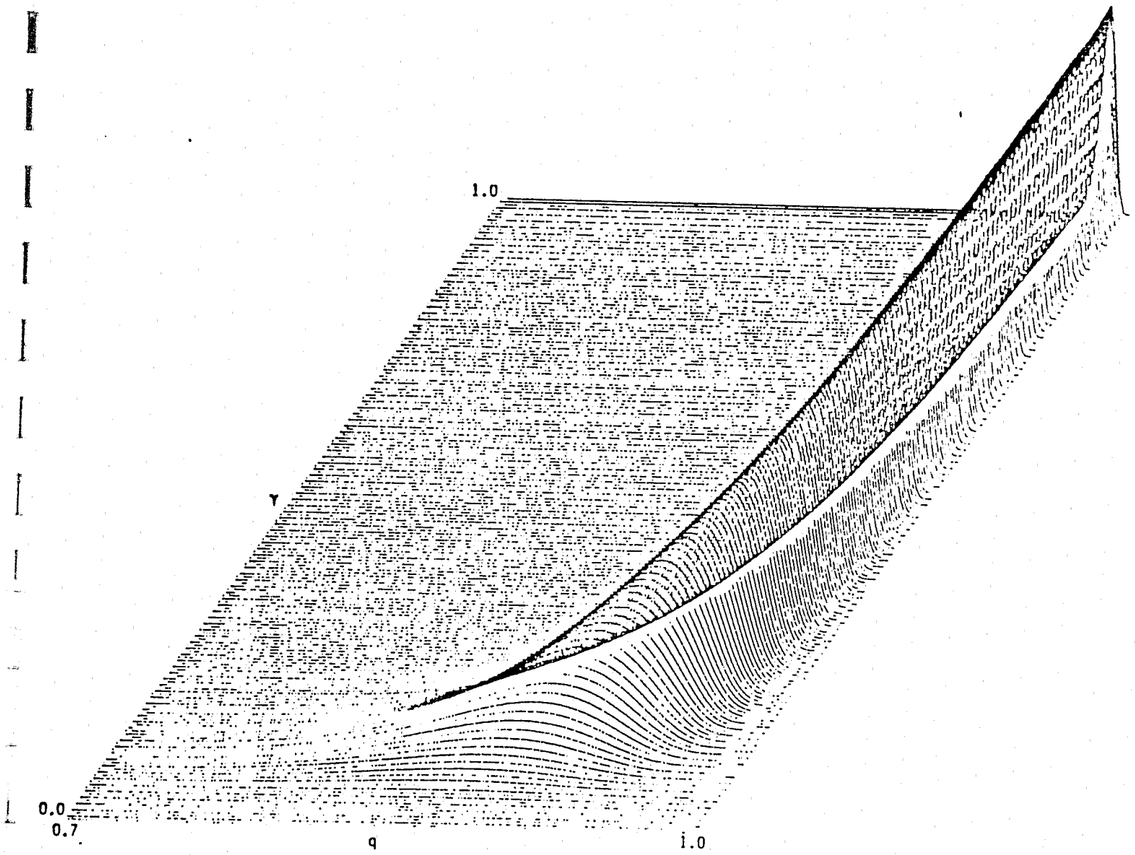
LIKELIHOOD FUNCTIONS USING UNIFORM PAROLE REPORTING DATA



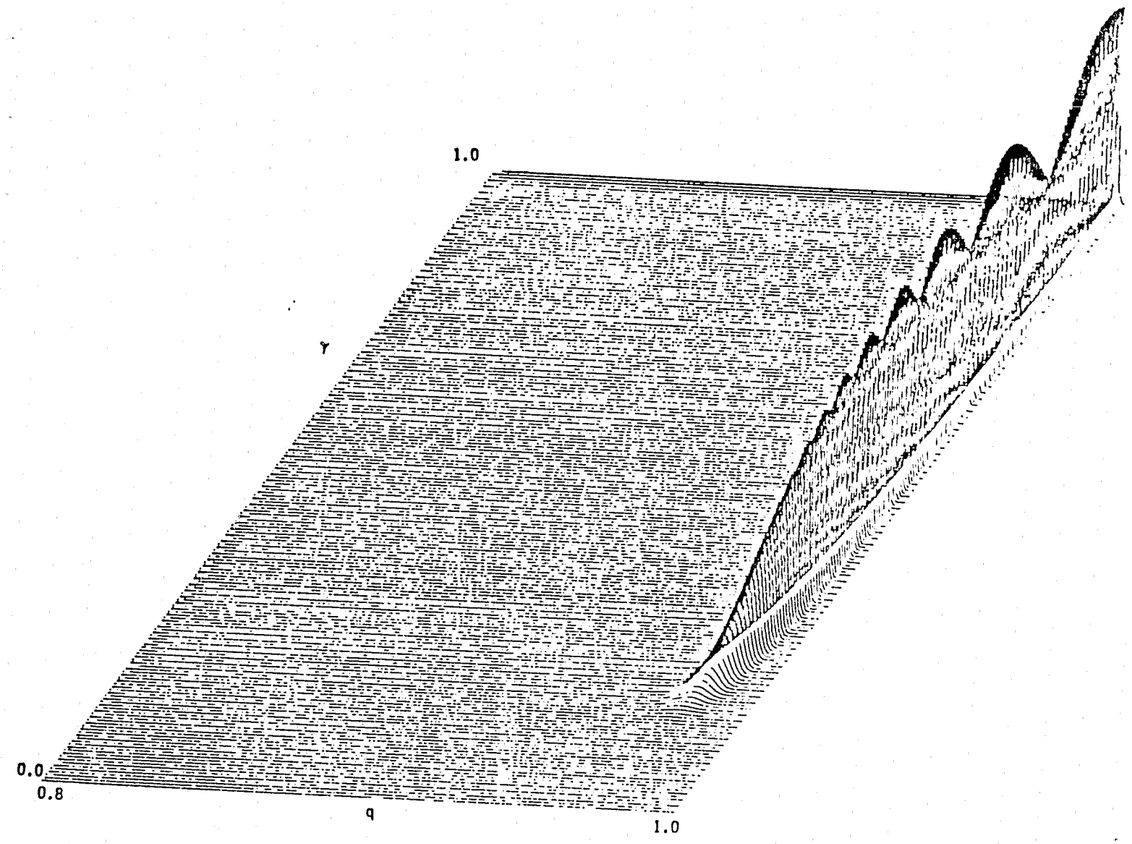
Likelihood Function for Alabama UPR Data



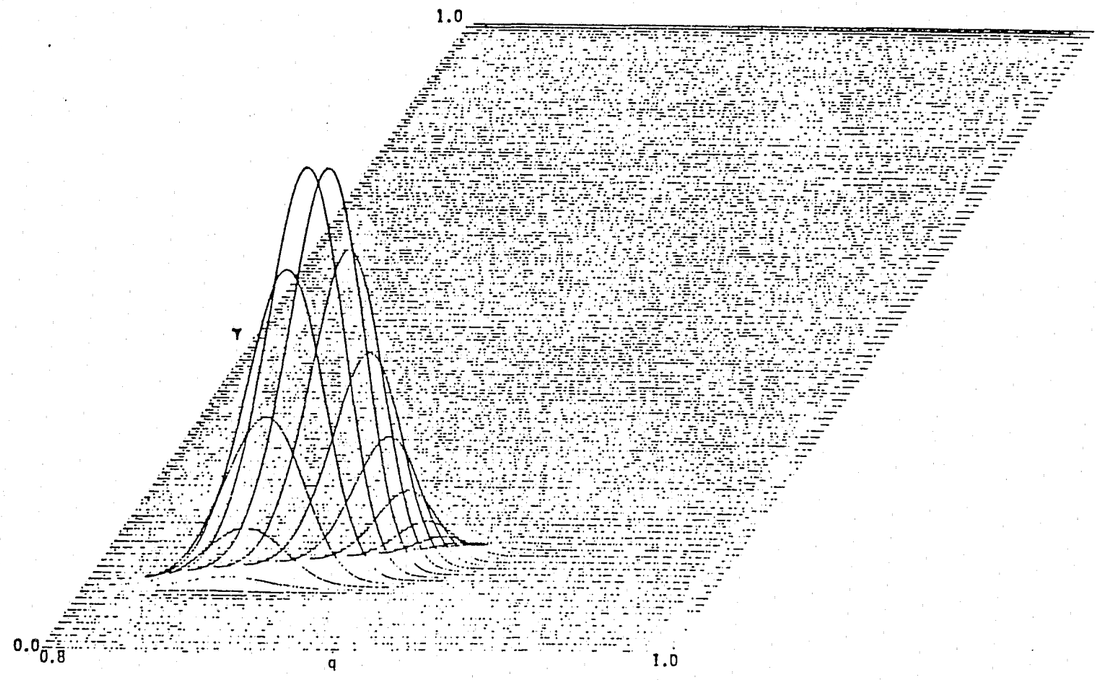
Likelihood Function for Alaska UPR Data



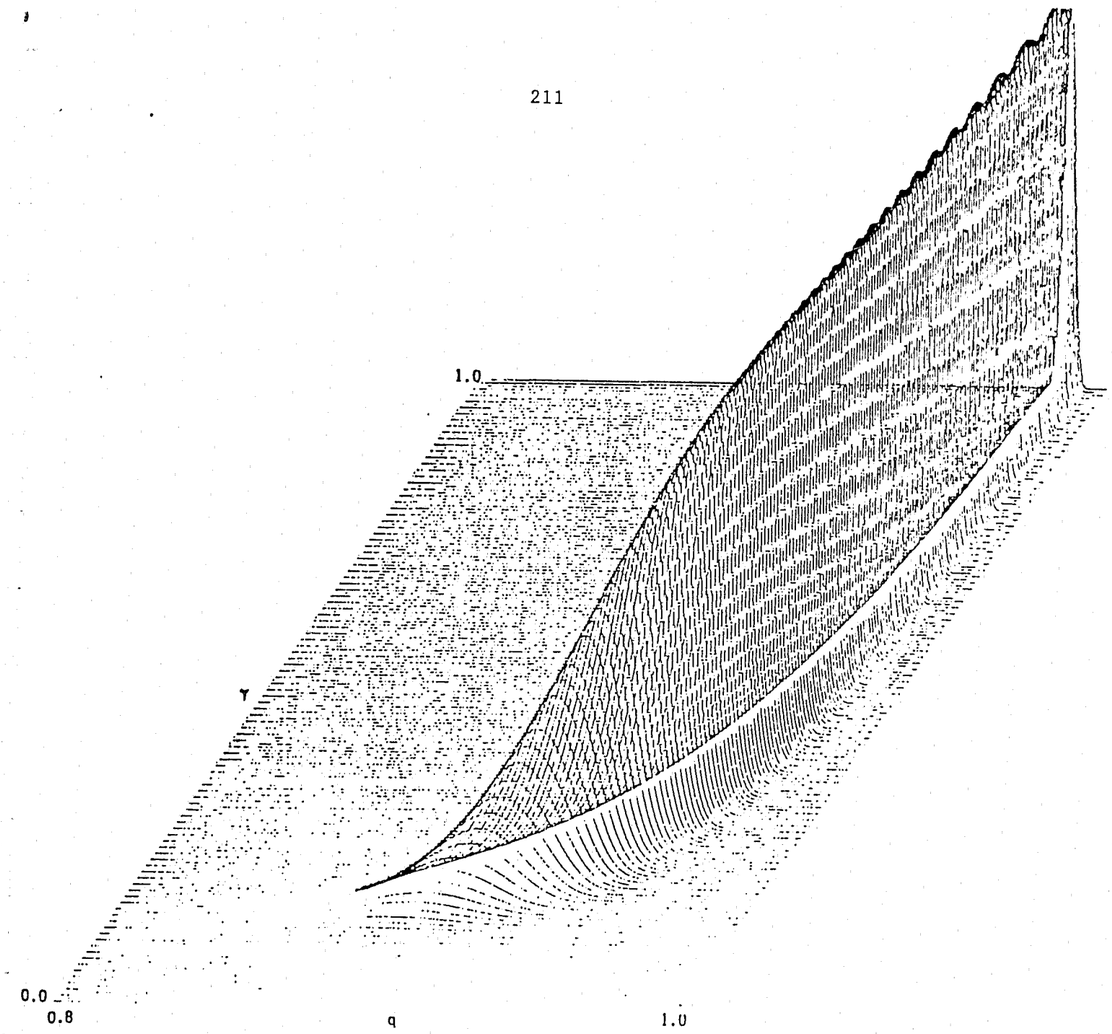
Likelihood Function for Delaware UPR Data



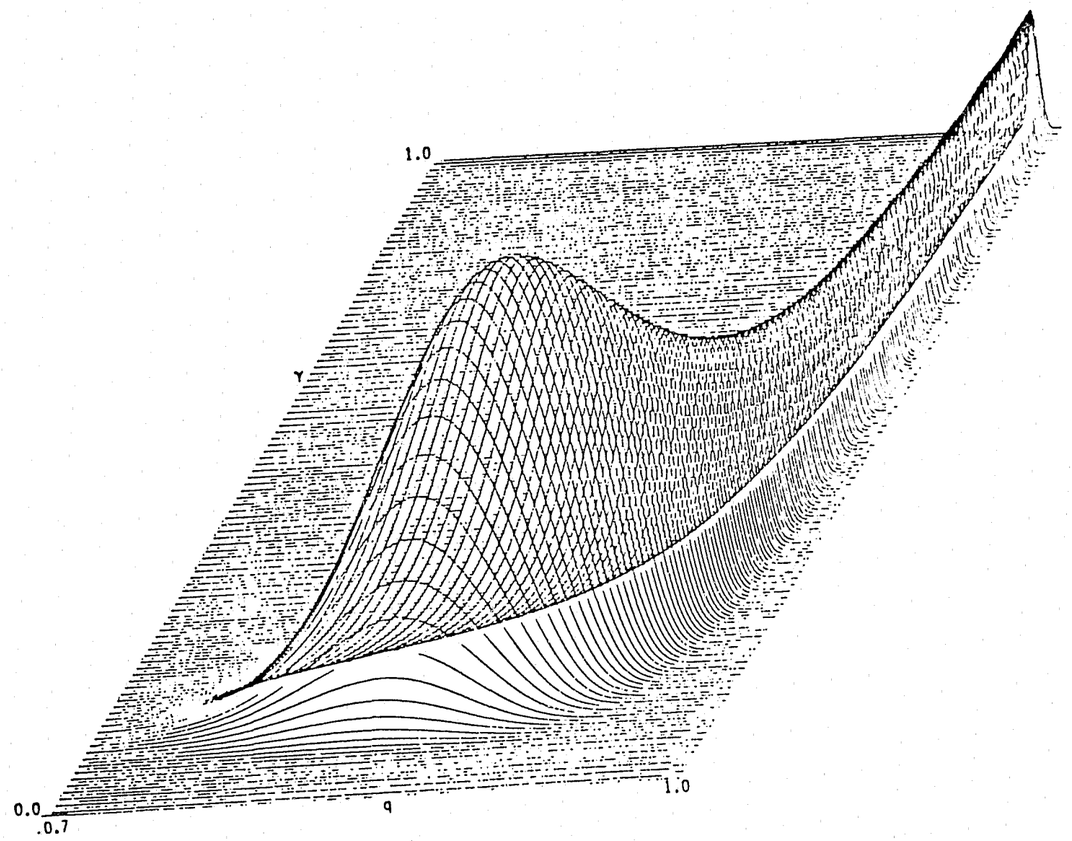
Likelihood Function for District of Columbia UPR Data



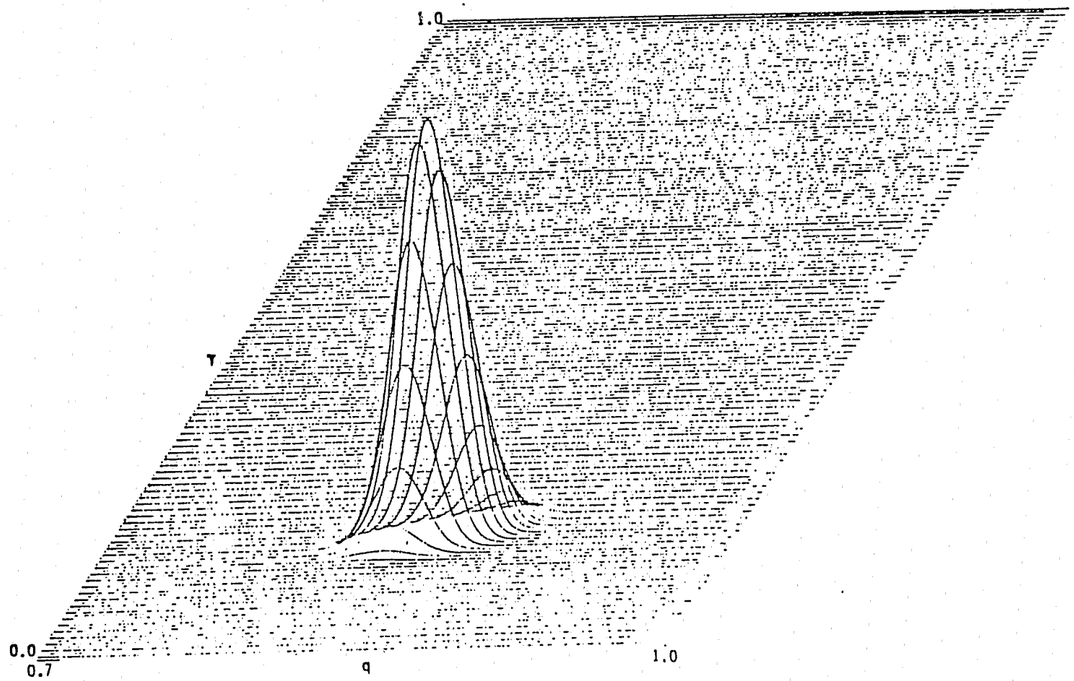
Likelihood Function for Florida UPR Data



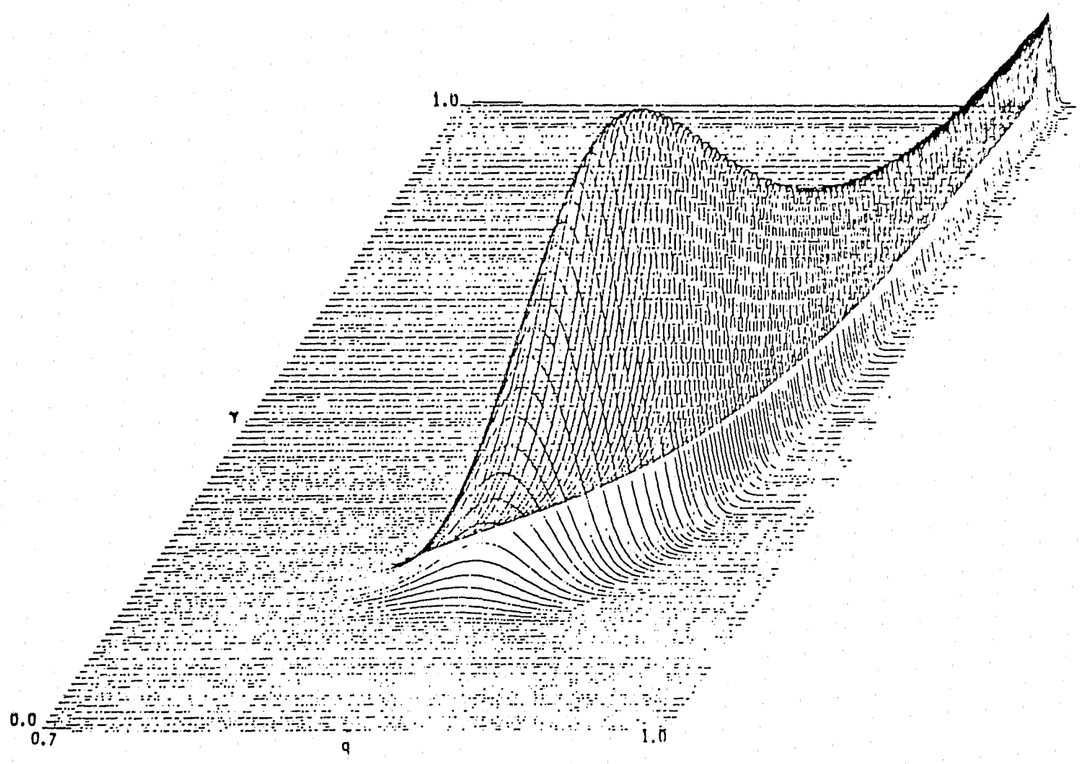
Likelihood Function for Georgia UPR Data



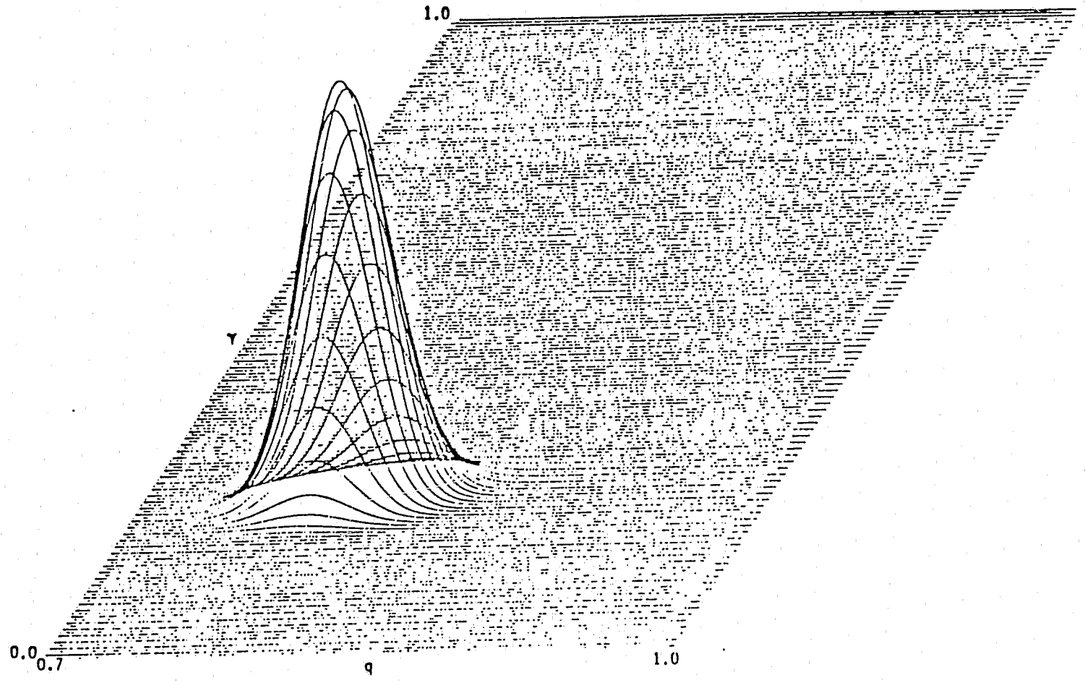
Likelihood Function for Idaho UPR Data



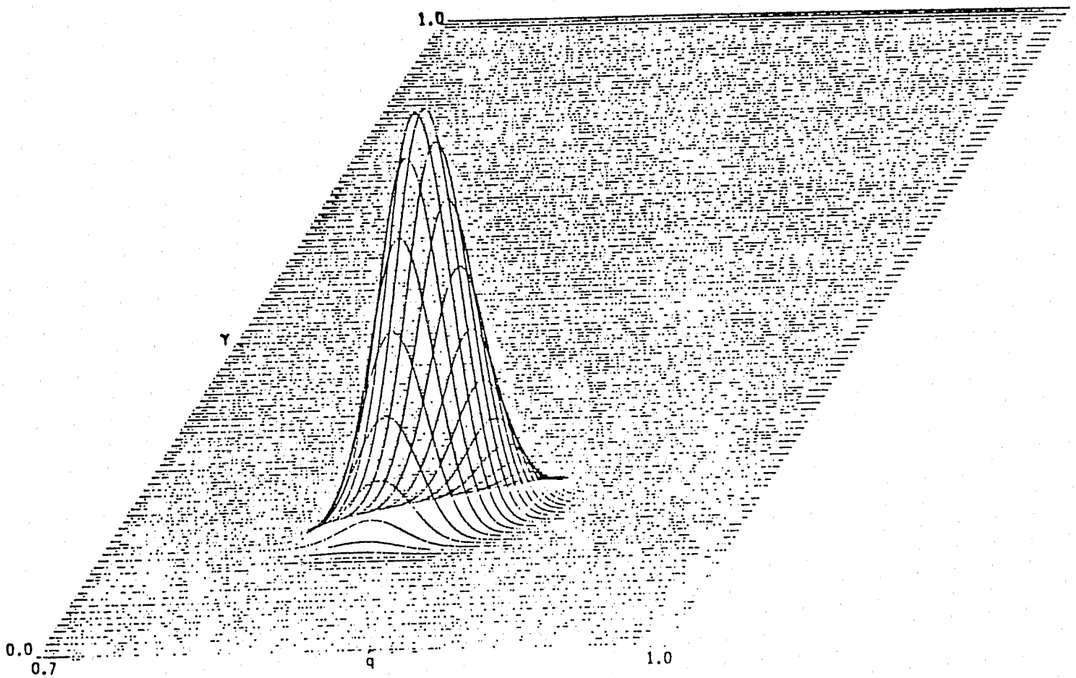
Likelihood Function for Illinois UPR Data



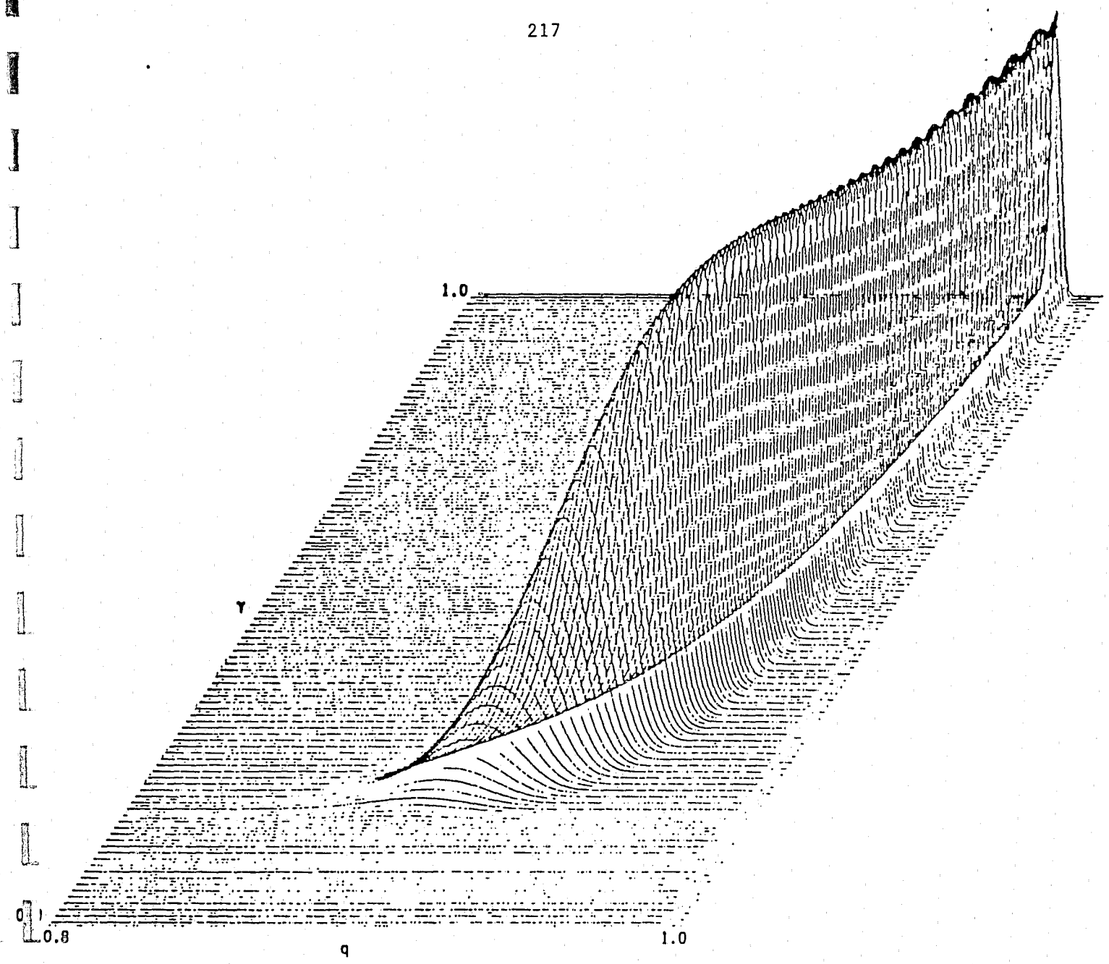
Likelihood Function for Iowa UPR Data



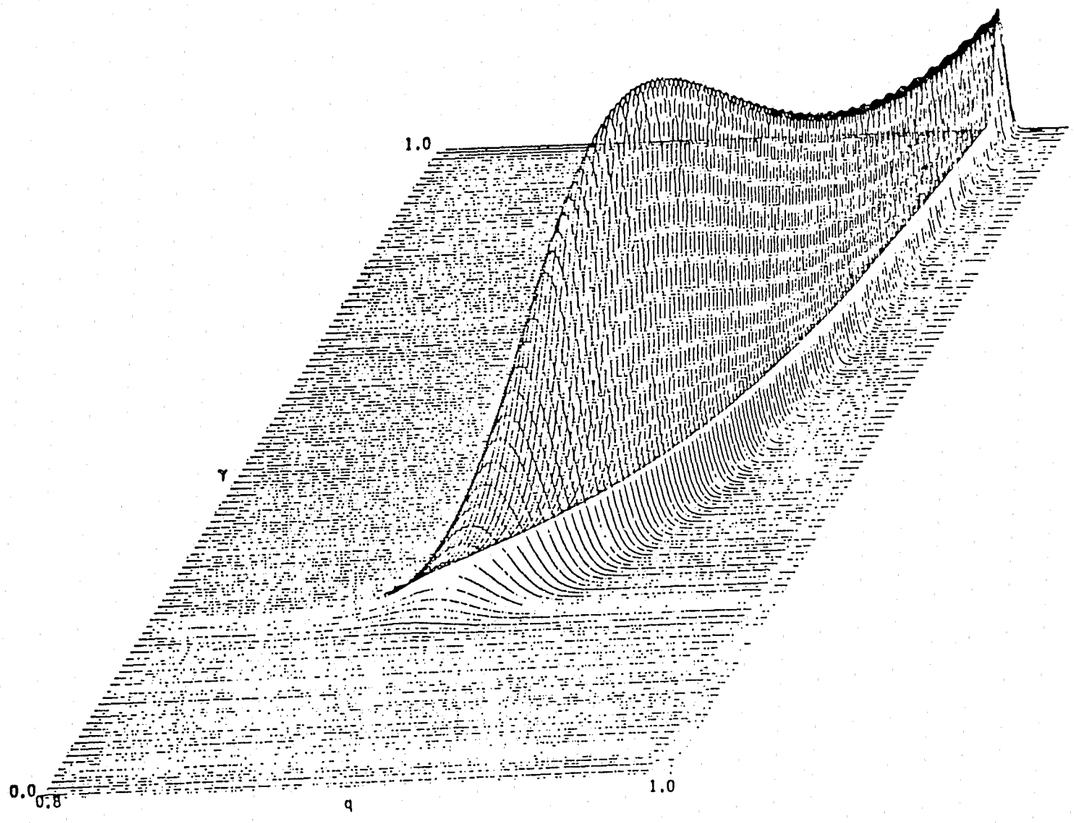
Likelihood Function for Kansas UPR Data



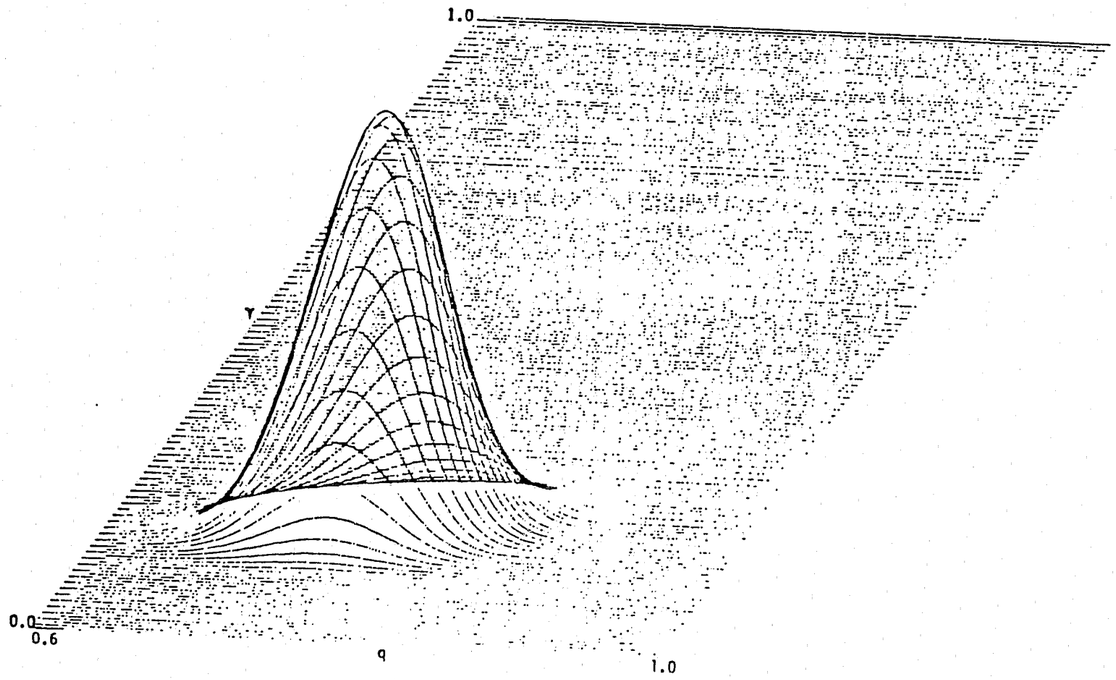
Likelihood Function for Kentucky UPR Data



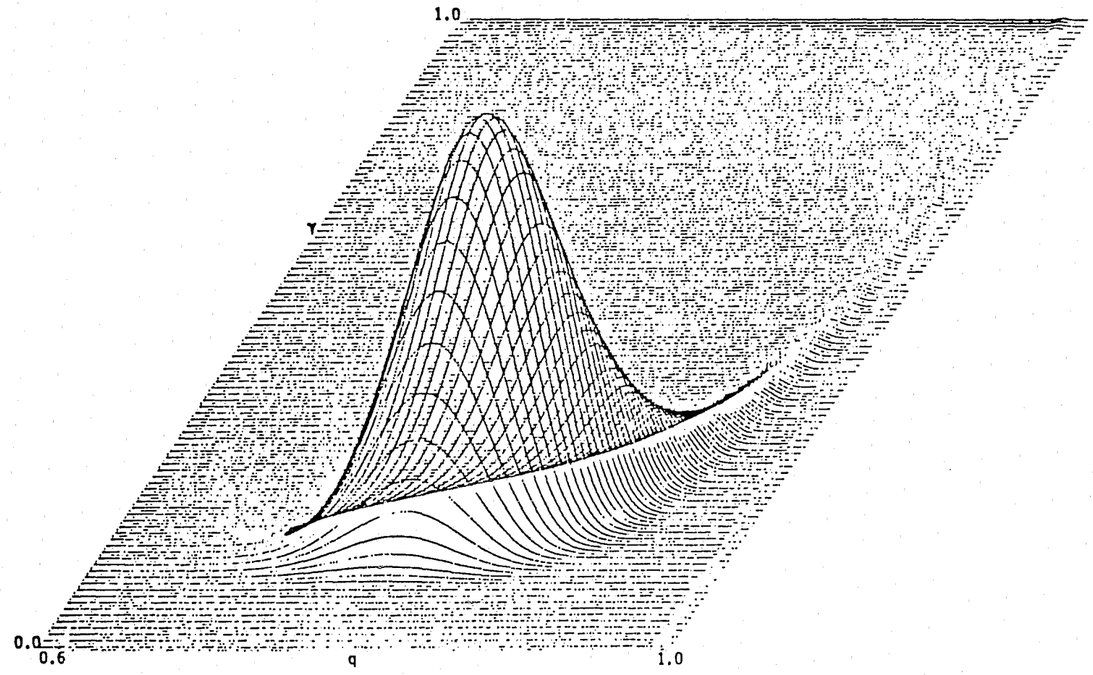
Likelihood Function for Maine UPR Data



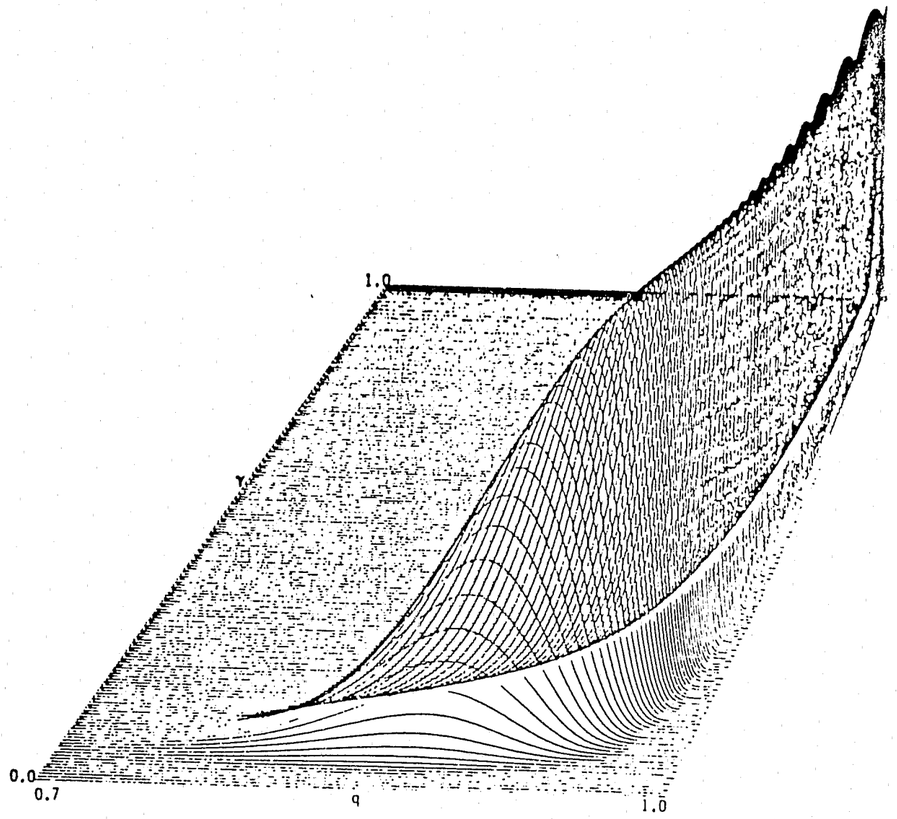
Likelihood Function for Michigan UPR Data



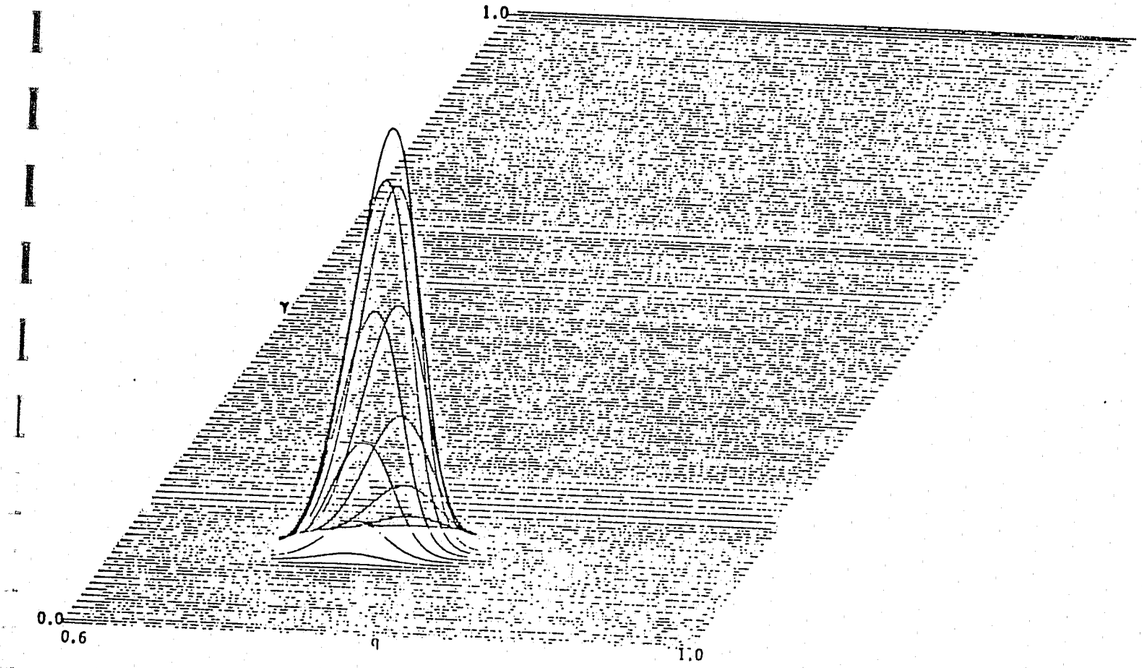
Likelihood Function for Missouri UPR Data



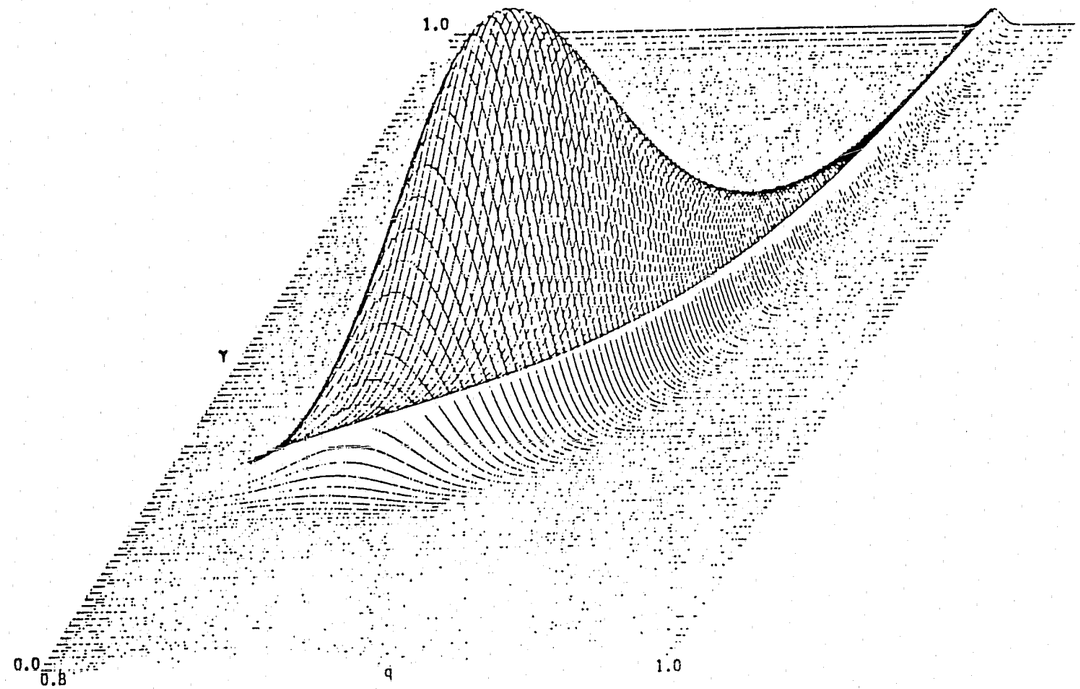
Likelihood Function for Nebraska UPR Data



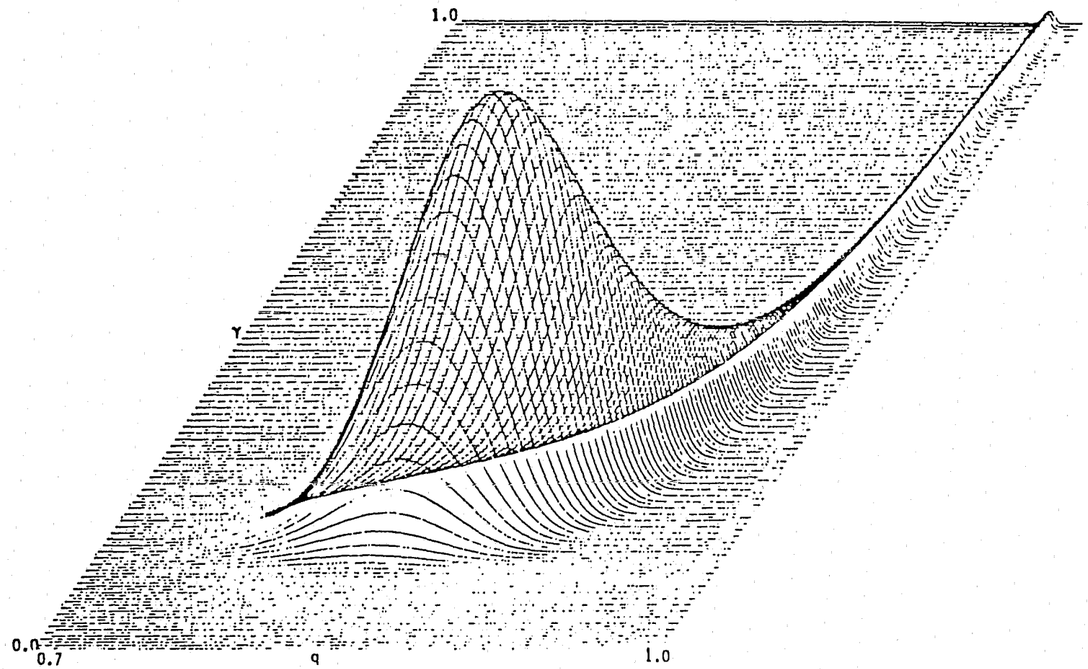
Likelihood Function for New Hampshire UPR Data



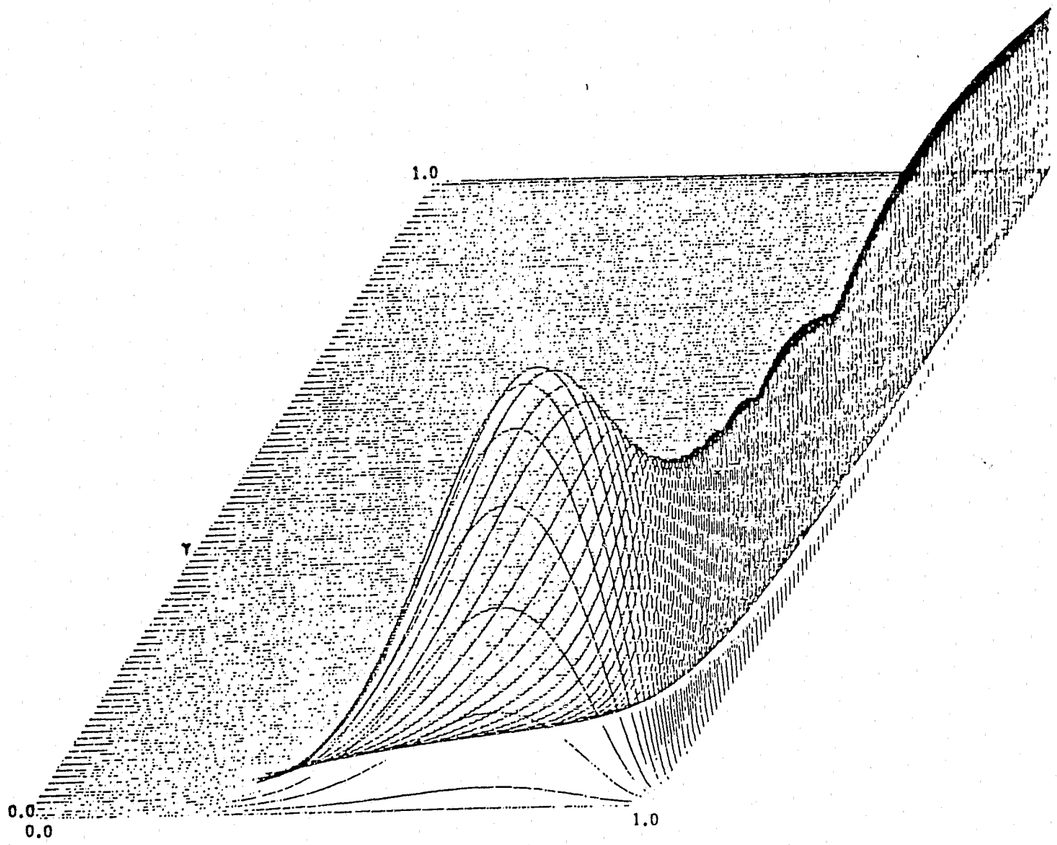
Likelihood Function for New Jersey UPR Data



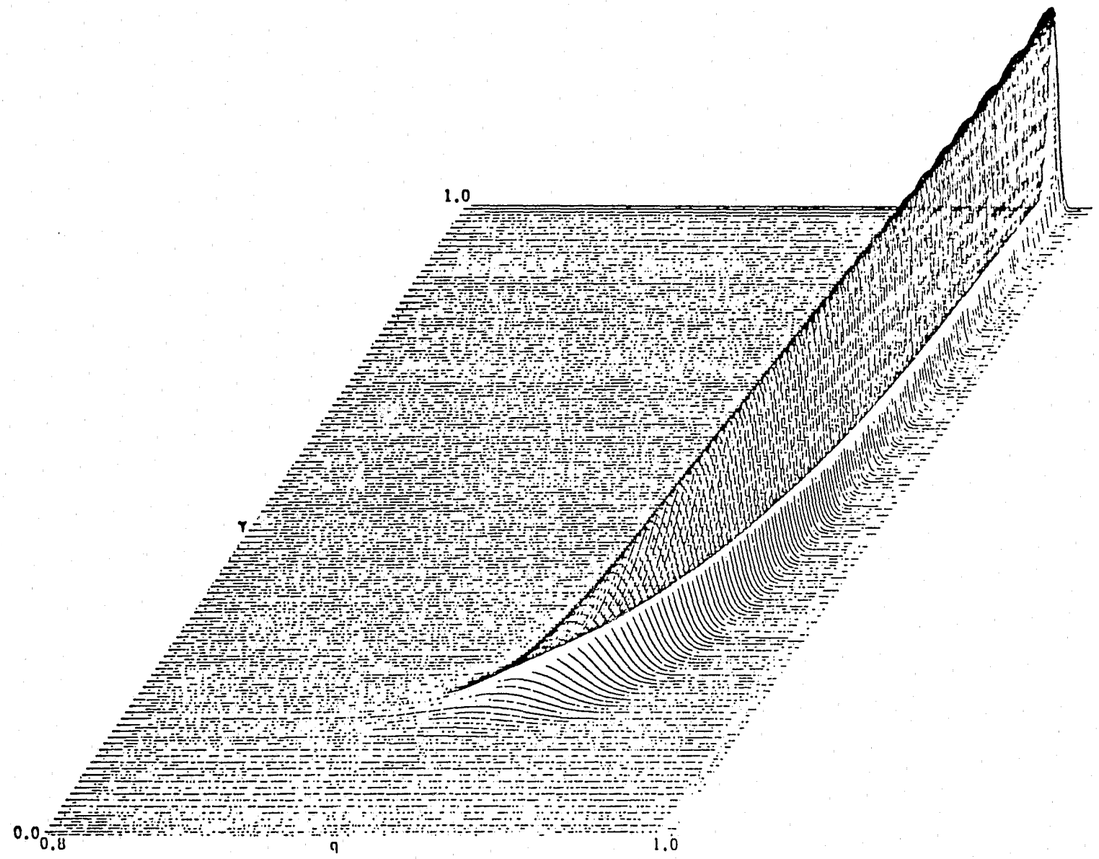
Likelihood Function for New Mexico UPR Data



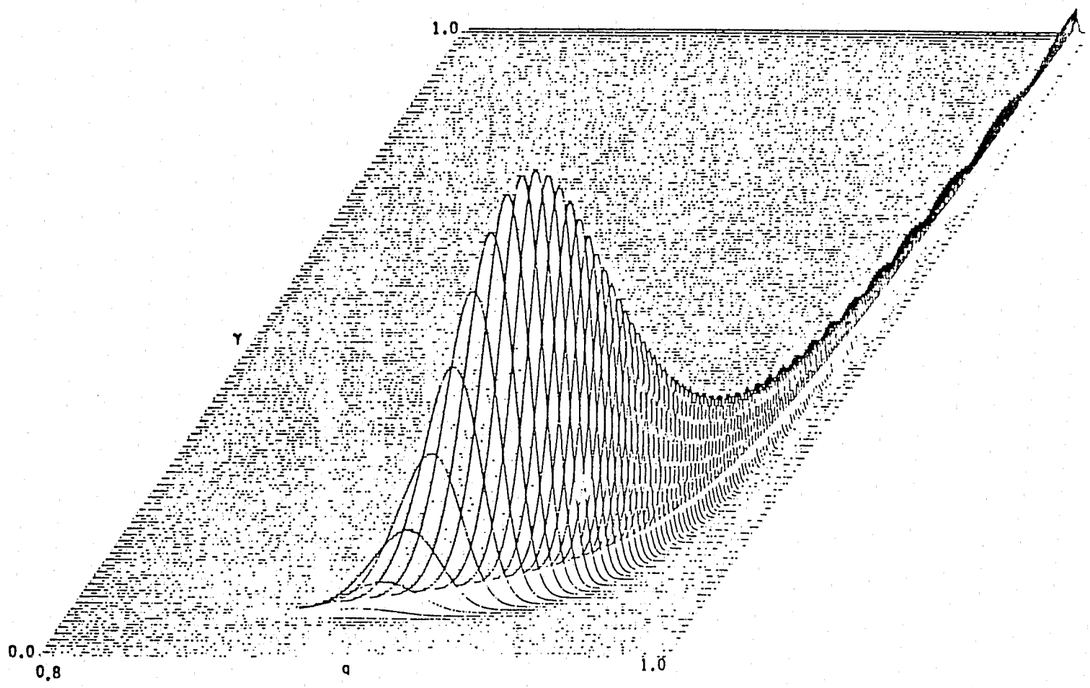
Likelihood Function for New York UPR Data



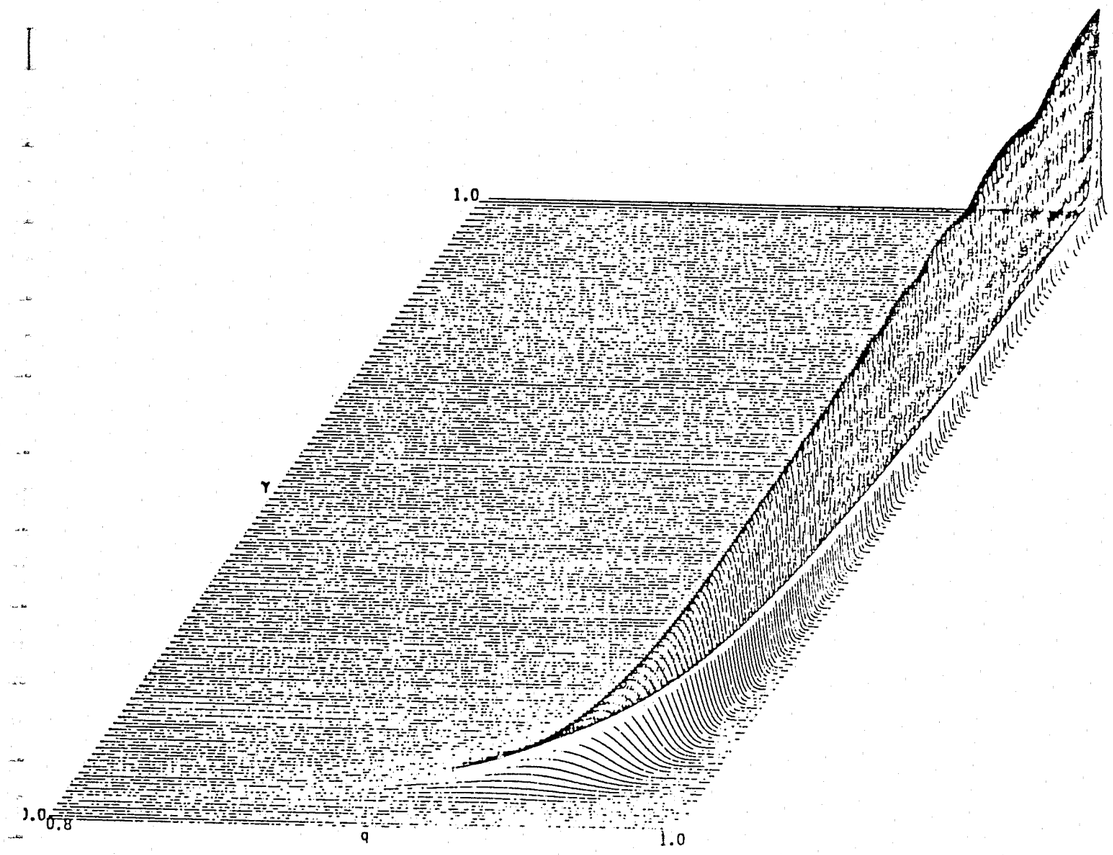
Likelihood Function for North Dakota UPR Data



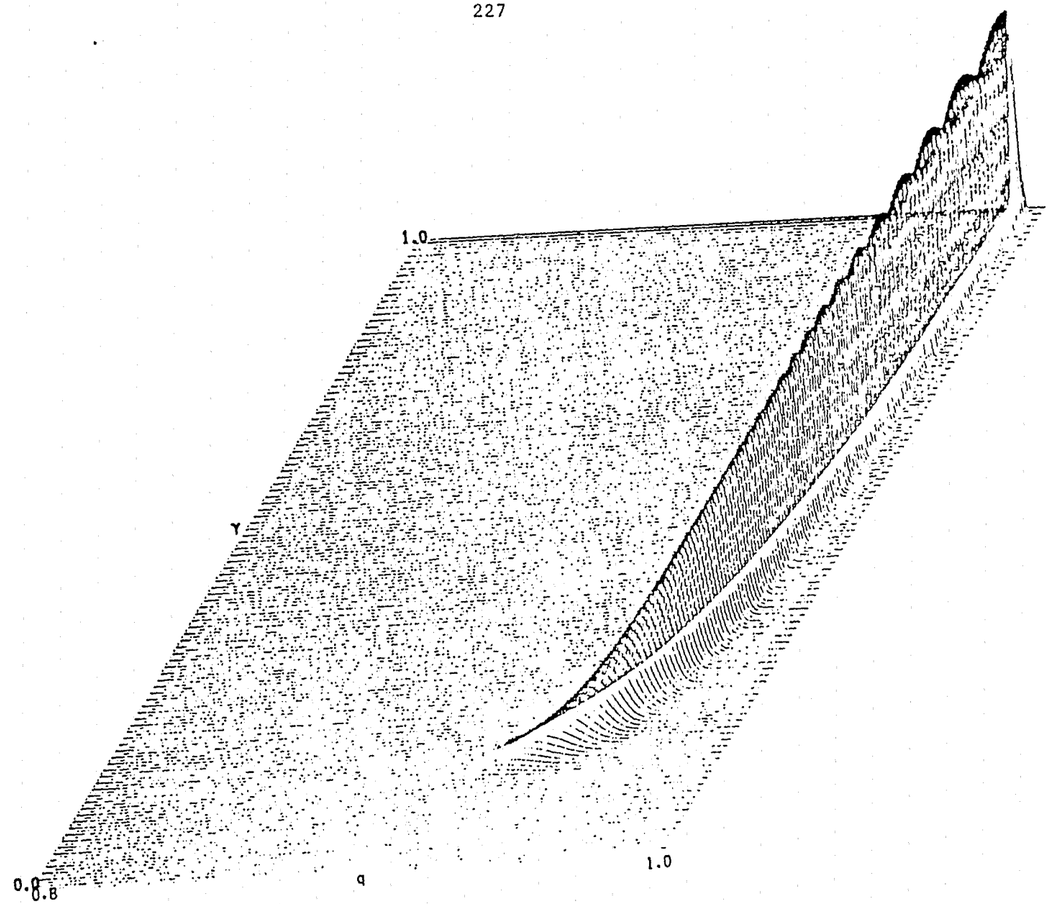
Likelihood Function for Ohio UPR Data



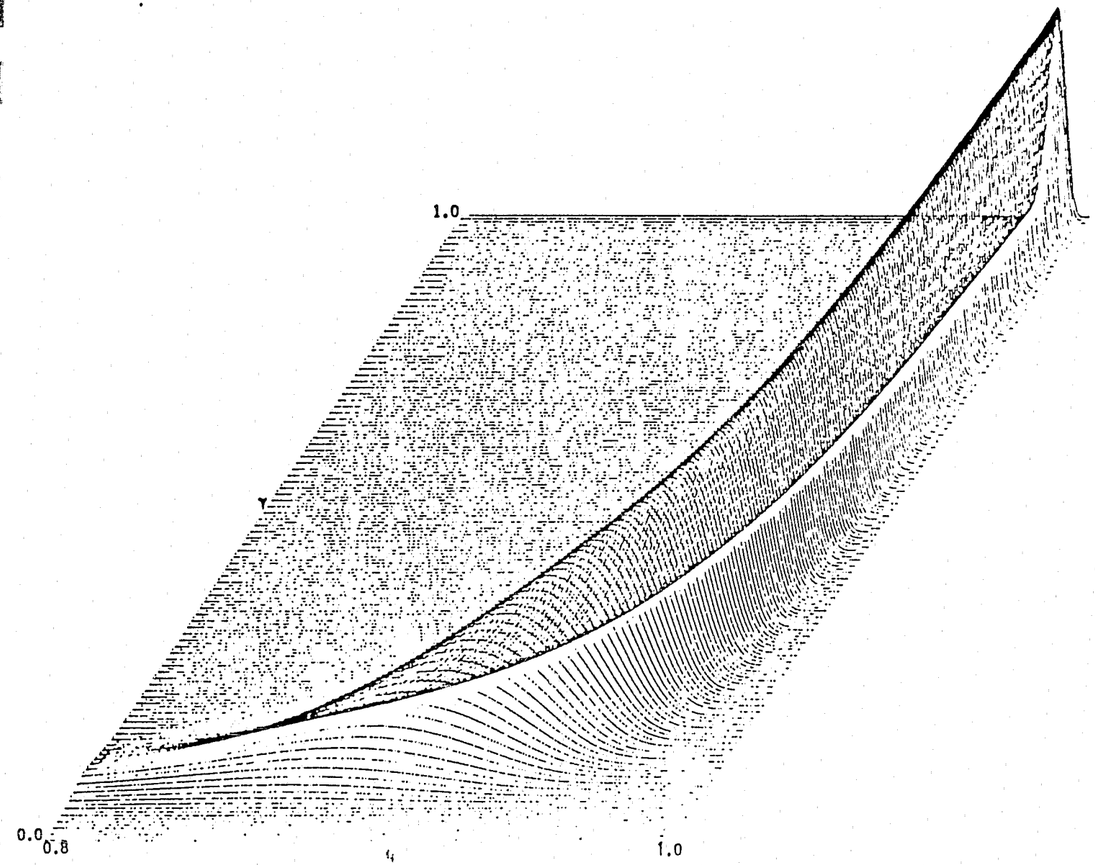
Likelihood Function for Pennsylvania UPR Data



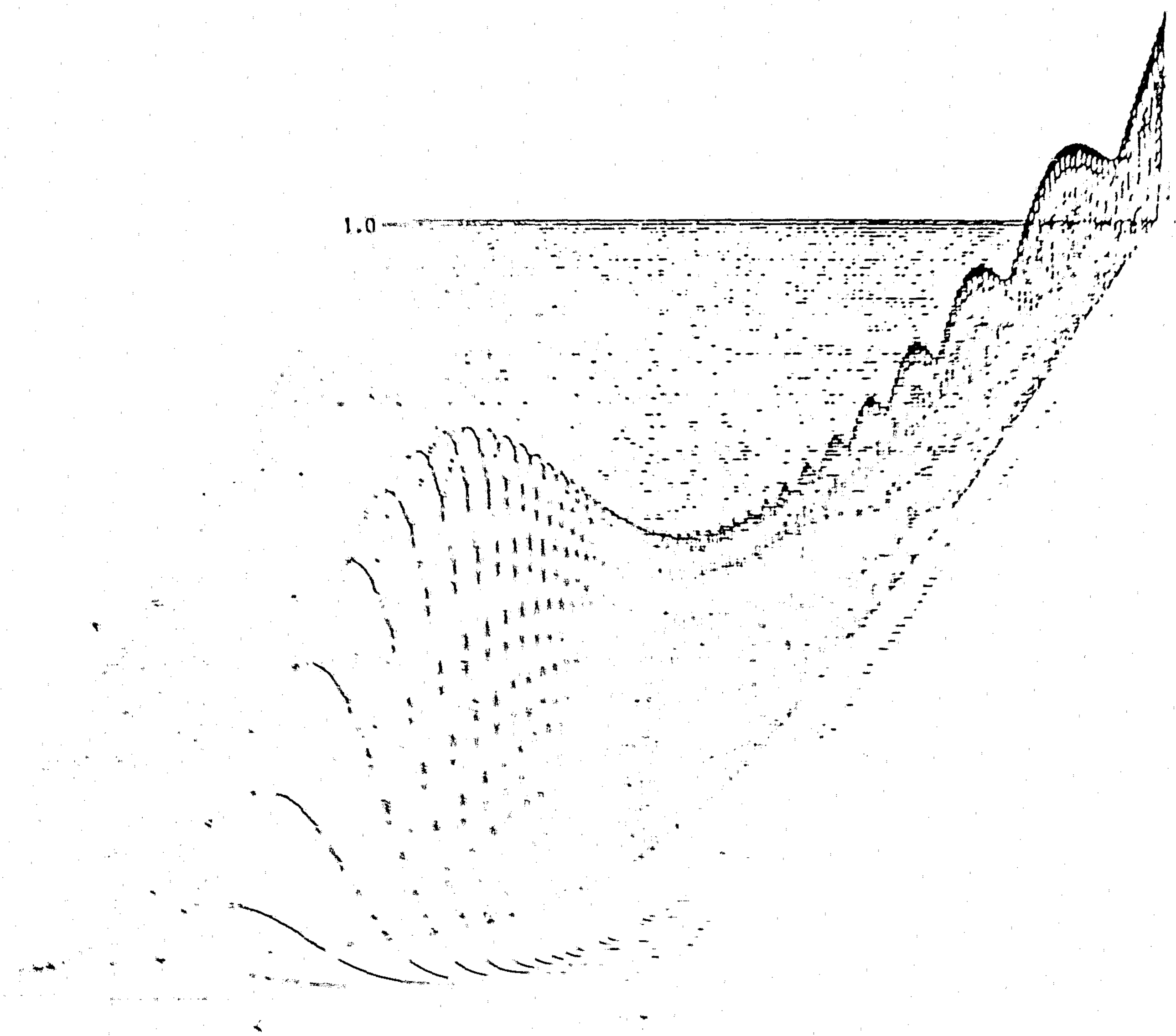
Likelihood Function for Puerto Rico UPR Data

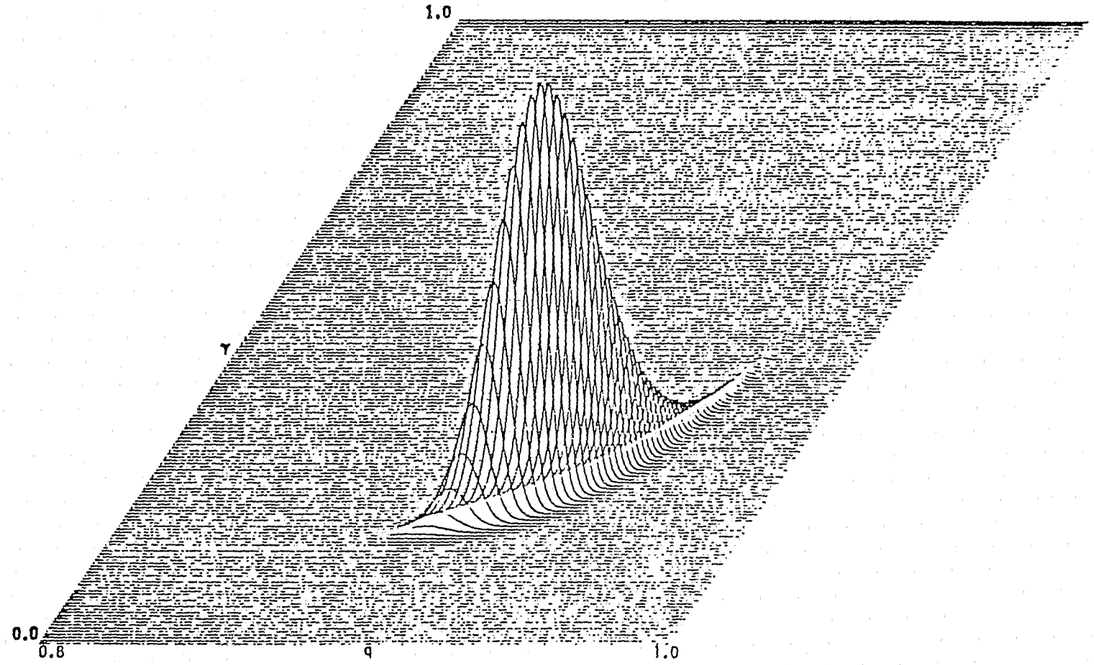


Likelihood Function for South Carolina UPR Data

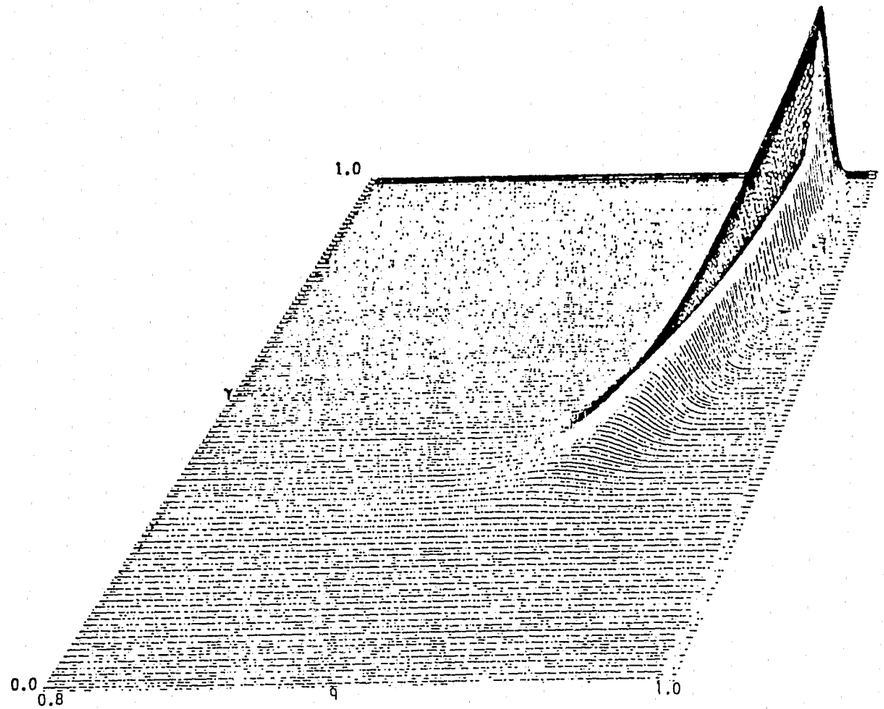


Likelihood Function for South Dakota UPR Data

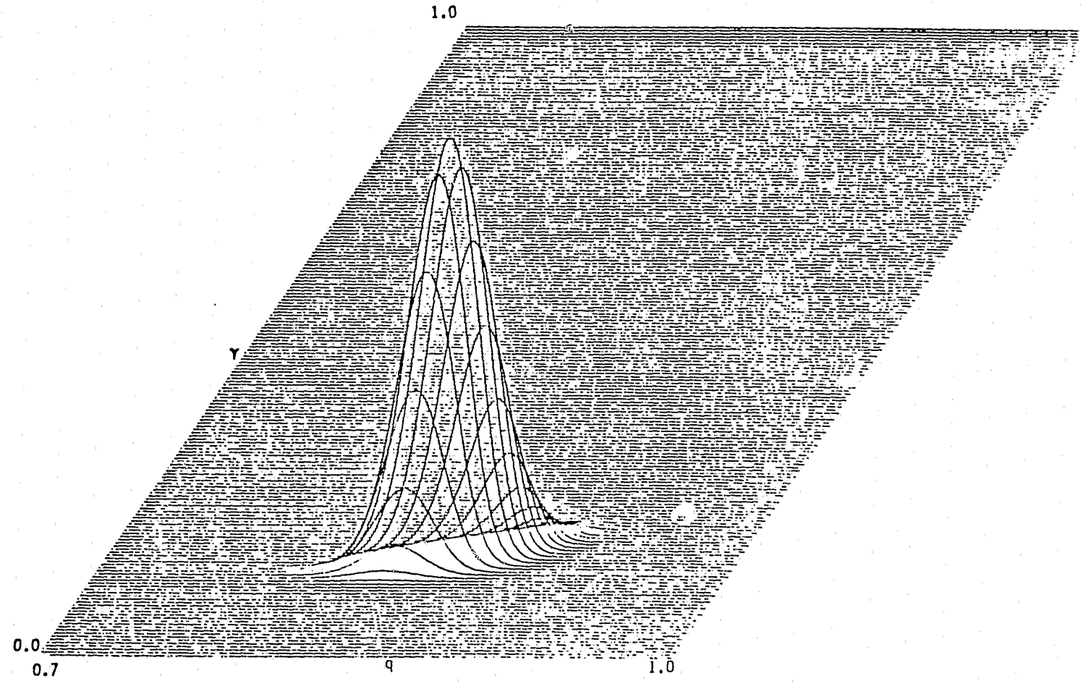

1.0



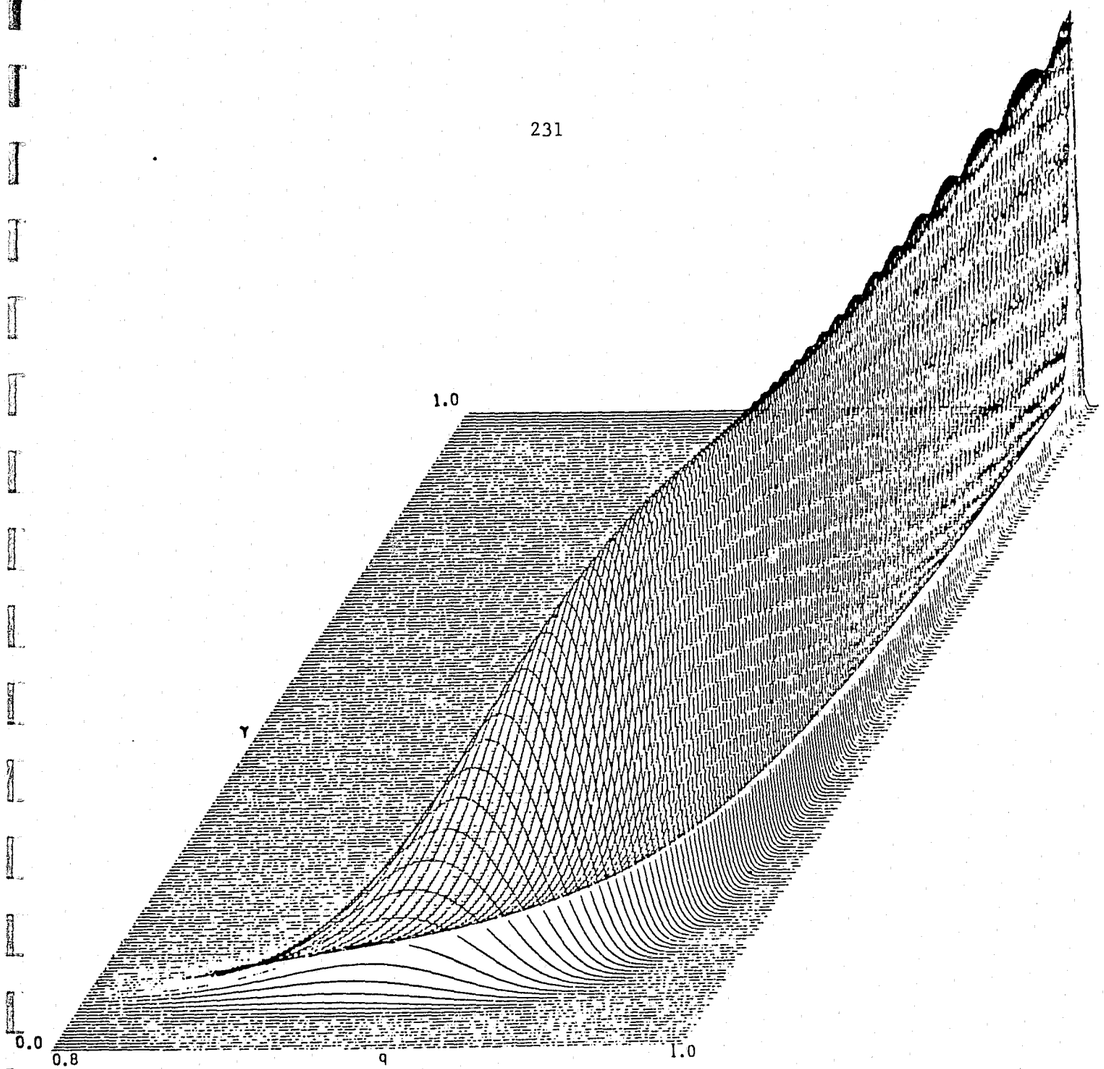
Likelihood Function for Texas UPR Data



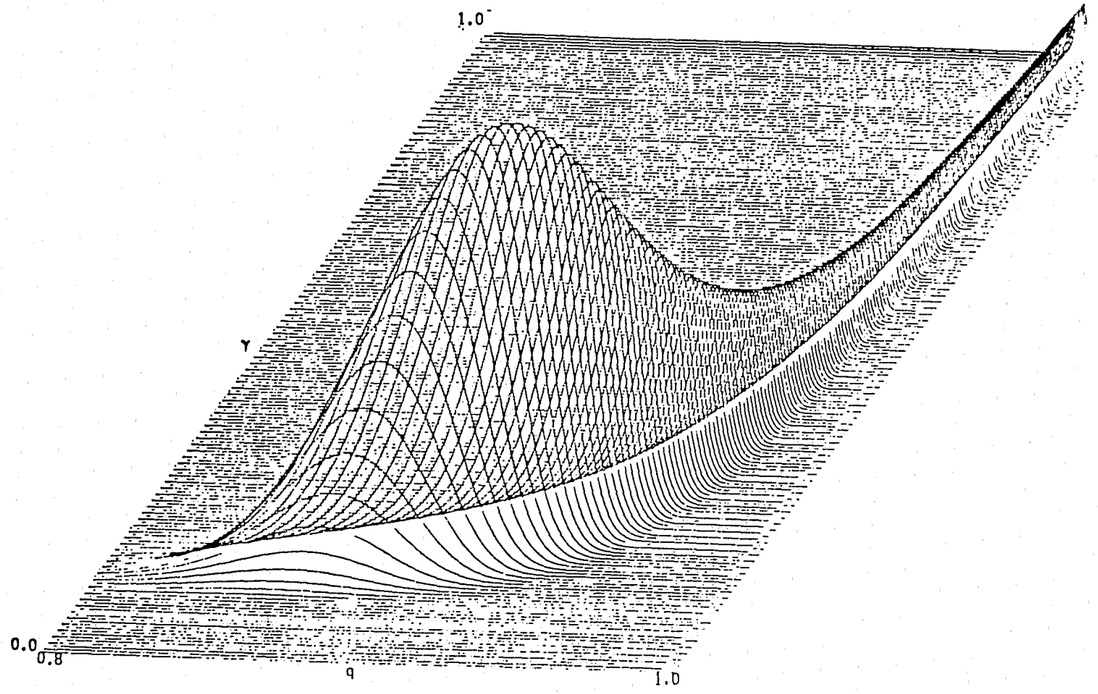
Likelihood Function for Utah UPR Data



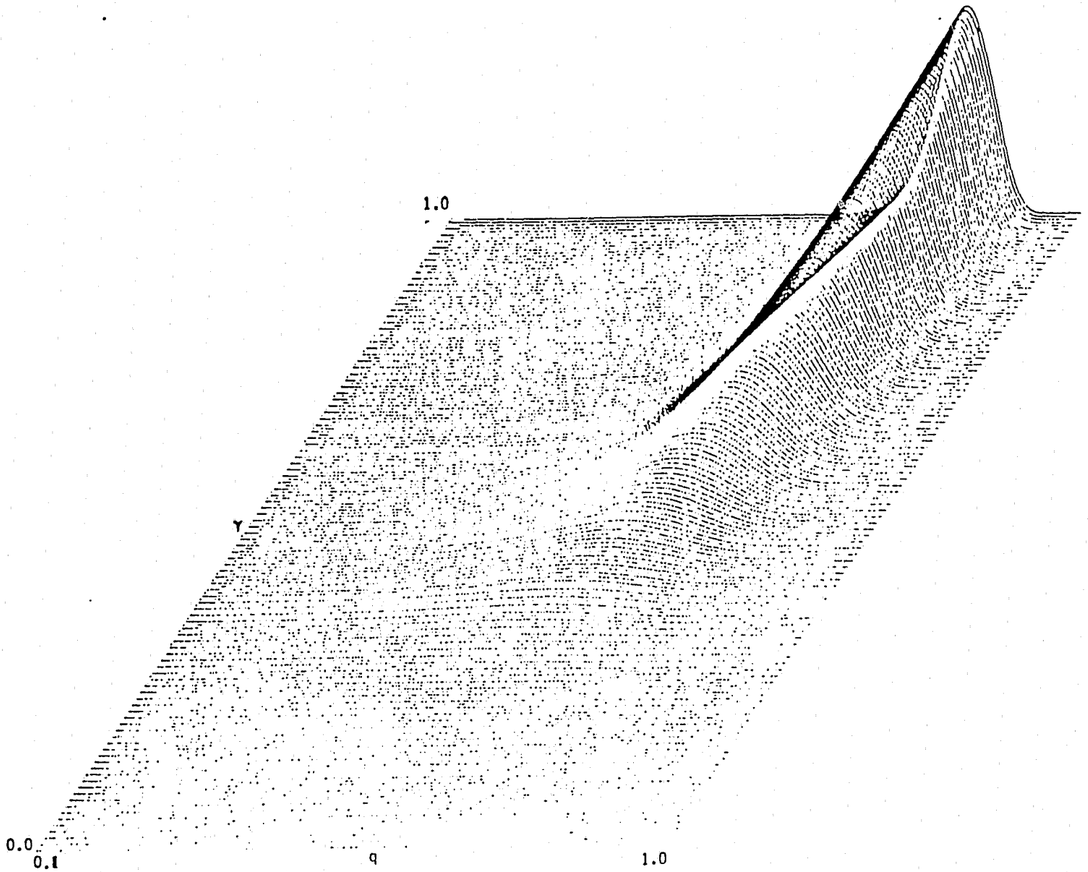
Likelihood Function for Virginia UPR Data



Likelihood Function for West Virginia UPR Data



Likelihood Function for Wisconsin UPR Data



Likelihood Function for Wyoming UPR Data

BIBLIOGRAPHY

(Entries marked with an asterisk * were used in the development of Table 6-1.)

Adams, R. and H.J. Vetter. "Probation Caseload Size and Recidivism Rate." British Journal of Criminology 11 (1971): 390-393.

*Adams, Stuart. "Effectiveness of the Youth Authority Special Treatment Program: First Interim Report." (1959). Cited in Lipton et al (1975).

*Adams, Stuart. "Effectiveness of Interview Therapy with Older Youth Authority Wards: An Interim Evaluation of the PICO Project." (1961) Cited in Lipton et al (1975).

*Adams, Stuart. "An Experimental Assessment of Group Counseling with Juvenile Probationers." (1964) Cited in Lipton et al (1975).

Alschuler, Albert. "Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for 'Fixed' and 'Presumptive' Sentencing." University of Pennsylvania Law Review. 126(1978): 550-577.

*Adamson, LaMay and Warren H. Dunham. "Clinical Treatment of Male Delinquents: A Case Study in Effort and Result." (1956) Cited in Lipton et al (1975).

American Friends Service Committee. Struggle for Justice. New York: Hill and Wang, 1971.

Anscombe, F.J. "Estimating a Mixed-Exponential Response Law." Journal of the American Statistical Association 56 (1961): 493-502.

Avi-Itzhak, B. and R. Shinnar. "Quantitative Models in Crime Control." Journal of Criminal Justice 1 (1973): 185-217.

Babst, Dean and John W. Mammering. "Probation versus Imprisonment for Similar Types of Offenders: A Comparison by Subsequent Violations." Journal of Research in Crime and Delinquency 2 (1965).

Baer. D.J., P.J. Jacobs, and F. E. Carr. "Instructors' Ratings of Delinquents After Outward Bound Survival Training and Their Subsequent Recidivism." Psychology Reports 36 (1975): 547-553.

Baldus, D. and J.W. Cole. "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment. Yale Law Journal 85 (1975): 170-186.

*Banks, Charlotte. "Reconviction of Young Offenders." (1964) Cited in Lipton et al (1975).

Barlow, Richard E., and Frank Proschan. Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart and Winston, Inc., 1975.

Beck, James and Peter Hoffman. "Time Served and Release Performance: A Research Note." Journal of Research in Crime and Delinquency 13 (1976): 127-132.

Belkin, Jacob, Alfred Blumstein, and William Glass. "Recidivism as a Feedback Process: An Analytical Model and Empirical Validation." Journal of Criminal Justice 1 (1973): 7-26.

Bennett, Lawrence and Max Ziegler. "Early Discharge: A Suggested Approach to Increased Efficiency in Parole." Federal Probation 39 (1975): 27-30.

*Benson, Sir George. "Prediction Methods and Young Prisoners." (1959) Cited in Lipton et al (1975).

Berkson, Joseph and Robert P. Gage. "Survival Curve for Cancer Patients Following Treatment." Journal of the American Statistical Association, 47 (1952): 501-515.

*Berntsen, Karen and Karl O. Christiansen. "A Resocialization Experiment with Short-Term Offenders." (1965) Cited in Lipton et al (1975).

Blalock, Hubert M. Jr. Social Statistics. New York: McGraw-Hill, 1972.

Bloom, H.S. "Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism." Sociological Methods & Research 8 (1979): 179-208.

Blumstein, Alfred, Jacqueline Cohen. "Estimation of Individual Crime Rates From Arrest Records." (1980, forthcoming) Journal of Criminal Law and Criminology.

Blumstein, Alfred and Jacqueline Cohen. "A Theory of the Stability of Punishment." The Journal of Criminal Law and Criminology 64 (1973): 198-207.

Blumstein, Alfred, Jacqueline Cohen, and Daniel Nagin, Editors. Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates. Washington, DC: National Academy of Sciences, 1978.

Blumstein, Alfred and Richard C. Larson. "Problems in Modeling and Measuring Recidivism." Journal of Research in Crime and Delinquency 8 (1971): 124-132.

Boag, John W. "Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy." Journal of the Royal Statistical Society 11 (1949): 15-53.

Boland, Barbara and J.Q. Wilson. "Age, Crime, and Punishment." The Public Interest 51 (Spring 1978): 22-34.

*Boston University, Law-Medicine Institute, Training Center in Youth Development. "Educational Counselors: Training for a New Definition of After-Care of Juvenile Parolees." (1966) Cited in Lipton et al (1975).

Bowers, W.J. and G.L. Pierce. "The Illusion of Deterrence in Issac Ehrlich's Research on Capital Punishment." Yale Law Journal 85 (1975): 187-208.

Box, George E.P. and George C. Tiao. Bayesian Inference in Statistical Analysis. Reading MA: Addison-Wesley Publication Company, 1973.

Brown, Lawrence D. "The Development of a Parole Classification System Using Discriminant Analysis." Journal of Research in Crime and Delinquency 24 (1978): 92-108.

*Burkhart, Walter R. and Arthur Sathmary. "An Evaluation of a Treatment Control Project for Narcotics Offenders: Phase 1 and 2," (1964). Cited in Lipton et al (1975).

Bury, Karl. Statistical Models in Applied Science. New York: John Wiley & Sons, 1975.

*California, Adult Authority, Divison of Adult Paroles. "Special Intensive Parole Unit, Phase 1: Fifteen Ma Caseload Study," (1956). Cited in Lipton et al (1975).

*California, Department of Corrections. "Intensive Treatment Program; Second Annual Report," (1958). Cited in Lipton et al (1975).

*California, Department of Corrections. "Special Intensive Parole Unit, Phase II: Thirty Man Caseload Study," (1958). Cited in Lipton et al (1975).

*Cambridge University, Department of Criminal Science. Dentention in Remand Homes, (1952). Cited in Lipton et al (1975).

Campbell, D.T. and A. Erlebacher. "How Regression Artifacts In Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." In J. Hellmeth, Ed., Compensatory Education: A National Debate, Vol. 3, Disadvantaged Child. New York: Brunner/Mazel, 1970.

Campbell, D.T. and H.L. Ross. "The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis." Law and Society Review 3 (1968): 33-53.

Carr-Hill, G.A. and R.A. Carr-Hill. "Reconviction as a Process." British Journal of Criminology 12 (1972): 35-43.

Chaiken, Jan M. "Estimates of Offender Characteristics Derived from the Rand Prison Inmate Survey." A working paper prepared for the Department of Justice. Rand Corporation, January, 1978.

Cirel, Paul, Patricia Evans, Daniel McGillis, and Debra Whitcomb. "An Exemplary Project: Community Crime Prevention in Seattle, Washington." National Institute of Law Enforcement and Criminal Justice, Washington, DC, September, 1977.

Coates,, Robert B., Alden D. Miller, and Lloyd E. Ohlin. Diversity in a Youth Correctional System: Handling Delinquents in Massachusetts. Cambridge, MA: Ballinger Publishing Co.

Cockerill, R. W. "Probation Effectiveness in Alberta." Canadian Journal of Criminology and Corrections 17 (1975): 284-291.

Cohen, Jacqueline. "The Incapacitative Effect of Imprisonment: A Critical Review of the Literature." In Blumstein et al., eds., Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates. Washington, DC: National Academy of Sciences, 1978, pp. 187-243.

Congress of the United States. Office of Technology Assessment. "A Preliminary Assessment of the National Crime Information Center and the Computerized Criminal History System." Washington, DC: US Government Printing Office, December 1978.

Cook, Thomas D. and Donald T. Campbell. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally College Publishing Company, 1979.

*Coombs, Keith A. "An Analysis of the Academic Educational Program in Wasington State Adult Correctional Institutions," (1965). Cited in Lipton et al (1975).

Corrections Digest. "LEAA/Commission 11 State Corrections Standards Project Proceeding on Schedule." Vol. 17, August 17, 1979.

Cox, D.R. and H.D. Miller. The Theory of Stochastic Processes. New York: Wiley and Sons, 1965.

Cox, George H. The Relative Impact of Georgia's Institutional Training Programs on the Post-Release Behavior of Adult Male Offenders. Atlanta: Georgia Department of Offender Rehabilitation, 1977.

Cutler, Sidney J. and Lillian M. Axtell. "Partitioning of a Patient Population with Respect to Different Mortality Risks." Journal of the American Statistical Association 58 (1963): 701-712.

Daniel, C. and F.S. Wood. Fitting Equations to Data. New York: Wiley 1971.

Davis, J.C. and A.J. Cropley. "Psychological Factors in Juvenile Delinquency." Canadian Journal in Behavioral Science 8 (1976): 68-77.

*Ditman, Keith S. and George Crawford. "The Use of Court Probation in the Management of the Alcohol Addict," (1965). Cited in Lipton et al (1975).

*Dombross, Earl E. and J.R. Silver. "Excerpt from Final Report of the San Francisco Rehabilitation Project for Offenders." (1966). Cited in Lipton et al (1975).

Deutsch, S.J., and F.B. Alt, "The Effect of the Massachusetts Gun Control Law on Gun-Related Crimes in the City of Boston," Evaluation Quarterly, 1, 4 (1977), 543-568.

Eck, John E. "Regulating Punishment: An Aggregate Population Dynamics Model of a Prison System." Paper presented at ORSA/TIMS Joint National Meeting. Milwaukee, October 15-17, 1979.

Ehrlich, I. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." Journal of Political Economy 81 (1973): 521-565.

Ehrlich, I. "Deterrence: Evidence and Inference." Yale Law Journal 85 (1975): 209-227.

Empey, Lamar T. and Maynard L. Erickson. The Provo Experiment. Lexington, MA: Lexington Books, 1972.

Empey, Lamar T. and Steven G. Lubeck. The Silverlake Experiment. Chicago: Aldine Publishing Co., 1971.

*Ericson, Richard C., et al. "The Application of Comprehensive Psycho-Social Vocational Services in Rehabilitation of Parolees...Period Covered by Report: January 1, 1965 through December 31, 1965," (1966). Cited in Lipton et al (1975).

Eysenck, S.B.G. and H.J. Eysenck. "Personality and Recidivism in Borstal Boys." British Journal of Criminology 14 (1974): 385-387.

*Feistman, Eugene G. "Comparative Analysis of the Willowbrook-Harbor Intensive Services Program, March 1, 1975 through February 28, 1966," (1966). Cited in Lipton et al (1975).

Feller, William. An Introduction to Probability Theory and its Applications. Volume I. 2nd Edition. New York: John Wiley & Sons, Inc., 1957.

Feller, William. An Introduction to Probability Theory and Its Applications. Volume II. New York: John Wiley & Sons, Inc., 1966.

Fienberg, Stephen, and Patricia Grambsch. "Appendix: An Assessment of the Accuracy of The Effectiveness of Correctional Treatment." Pp. 119-147 in Sechrest et al (1979).

Fishman, Robert. Criminal Recidivism in New York City: An Evaluation of the Impact of Rehabilitation and Diversion Services. New York: Praeger Publishers, 1977.

Fogel, David. We Are the Living Proof. Cincinnati: W.H. Anderson Publishing Co., 1979.

Forst, B. "Participation in Illegitimate Activities: Further Empirical Findings." Policy Analysis 2 (1976): 477-492.

*Fox, Vernon. "Michigan's Experiment in Minimum Security Penology." (1950). Cited in Lipton et al (1975).

*Freeman, Howard E. and H. Ashley Weeks. "Analysis of a Program of Treatmentof Delinquent Boys." (1956). Cited in Lipton et al (1975).

*Garrity, Donald Lee. "The Effects of Length of Incarceration upon Parole Adjustment and Estimation of Optimun Sentence: Washington State Correctiona Institutions," (1956). Cited in Lipton et al (1975).

*Gearhart, J. Walter, Harold L. Keith, and Gloria Clemmons. "An Anaysis of the Vocationa Training Program in the Washington State Adult Correctional Institutions," (1967) Cited in Lipton et al (1975).

Golomb, Henry D. and Howard M. Bunch. "Stochastic Analysis of Future Vehicle Populations." Report for the Highway Safety Research Institute, The University of Michigan, Ann Arbor, April, 1978.

Goodman, Leo A. "Statistical Methods for the Mover-Stayer Model." Journal of the American Statistical Association 56 (1961): 841-868.

Gordon, Andrew C., David McDowall, Michael Maltz, and Richard McCleary. "Evaluating a Delinquency Intervention Program: A Comment." Criminal Justice Newsletter 9 (1978): 1-4.

Gottfredson, D.M. and K.B. Ballard. "Differences in Parole Decisions Associated with Decision-Makers." Journal of Research in Crime and Delinquency. 3 (1966): 112-119.

*Great Britain, Home Office. The Sentence of the Court: A H Handbook for Courts on the Treatment of Offenders, (1964). Cited in Lipton et al (1975). Cited in Lipton et al (1975).

Greenberg,David F. "The Correctional Effects of Corrections: A Survey of Evaluations." In D.F. Greenberg, Ed., Corrections and Punishment. Beverly Hills: Sage Publications, 1977.

Greenberg, David F. "Recidivism as Radioactive Decay." Journal of Research in Crime and Delinquency 24 (1978): 124-125.

Greenberg, David F. "The Incapacitative Effect of Imprisonment: Some Estimates." Law and Society Review 9 (1975): 541-580.

Greenberg, Richard A., Steven Bayard, and David Byar. "Selecting Concomitant Variables Using a Likelihood Ratio Step-Down Procedure and a Method of Testing Goodness of Fit in an Exponential Survival Model." Biometrics 30 (1974): 601-608.

Greenwood, Peter W., "Rand Research on Criminal Careers: Progress to Date," Report No. N-1286-DOJ, Rand Corporation, Santa Monica, California, August, 1979.

Griswold, David B. "A Comparison of Recidivism Measures." Journal of Criminal Justice 6 (1978): 247-252.

Grizzle, Gloria A. "A Typology of Performance Measures for Correctional Programs." Working paper 79-1, March, 1979. "Criteria for Rating Corrections Performance Measures." Working paper 79-2, April, 1979. "A Conceptual Framework for Correctional Performance Measurement." Working paper 79-3, May, 1979. The Osprey Company, Raleigh, NC. Prepared under grant 78-NI-AX-0130 from the National Institute of Law Enforcement and Criminal Justice.

Gross, Alan J. and Virginia A. Clark. Survival Distributions: Reliability Applications in the Biomedial Sciences. New York: John Wiley & Sons, 1975.

Gulland, J.A. "On the Estimation of Population Parameters from Marked Members." Biometrika 43 (1955): 269-270.

*Guttmam, Evelyn S. "MMPI Measured Chages in Treated and Untreated Youth Authority Wards Judged in Need of Psychiatric Treatment." (1961). Cited in Lipton et al (1975).

*Guttman, Evelyn S. "Effects of Short-Term Psychiatric Treatment on Boys in Two California Youth Authority Institutions," (1963). Cited in Lipton et al (1975).

Haight, Frank A. Handbook of the Poisson Distribution. New York: John Wiley and Sons, Inc., 1967.

Center for Research in Criminal Justice

University of Illinois at Chicago Circle
Box 4348 Chicago 60680

ON RECIDIVISM:
Exploring Its Properties as a Measure of Correctional Effectiveness

by

Michael D. Maltz
Center for Research in Criminal Justice
University of Illinois at Chicago Circle
Chicago, Illinois 60680

January 1981

Final Report on Grant Number 77-NI99-0073 from the National Institute of Law Enforcement and Criminal Justice. Points of view or opinions stated in this report are those of the author and do not necessarily reflect the official position or policies of the U.S. Department of Justice. Computing services were provided by the UICC Computer Center; their assistance is gratefully acknowledged.

ACKNOWLEDEGEMENTS

This report is the outgrowth of a research project funded by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. This research is not the result of one person's efforts. The contribution of other participants has been major and significant. Simply to reel off a list of names would ignore the part they played in shaping the result.

The idea for the research grew out of discussions between Richard McCleary and me on measures of correctional effectiveness, discussions which culminated in a paper (Maltz & McCleary, 1977) and in the grant application. McCleary and I were co-prinicpal investigators on the project for the first year, until he left the University of Illinois at Chicago Circle for the warmer climes of Tempe and Arizona State University. Many of the ideas in this report stem from our discussions.

Stephen Pollock was the primary consultant on the quantitative aspects of this research. His ability to understand the nature of the process under consideration and to develop useful mathematical models of the process exemplifies the very essence of operations research. So too with Carl Harris, who consulted on this grant in its later stages; in fact, it was the seminal paper written by him and Stephen Stollmack (1974) that led directly to the methods we report on herein.

Frank Merritt of the University of Toledo Law School did a masterful job of analyzing the statutes, case law and administrative regulations pertaining to parole revocation. His paper (Merritt, 1980) formed the basis for our discussion of parole organization in Chapter 5. Other aspects of the organization and structure of correctional agencies, and of the different procedures used to assess correctional effectiveness, were the responsibility of research assistants Louise Cainkar, Laura Frazier and Steven Mihajlovic. Steve was also responsible for much of the computer and graphics work in this report. And Mary Pallen somehow was able to keep the administrative and budgetary side of the project running smoothly; she also doubled as research assistant upon occasion.

The project was fortunate in having an advisory board that worked effectively with us. It consisted of: Melvin Axilbund, former Director of the American Bar Association Commission on Correctional Facilities and Services; John Conrad, now with the American Justice Institute; John Henning, Research and Evaluation Manager of the Illinois Department of Corrections; and Gerald Strathman, Director of Research, Minnesota Department of Corrections. They were instrumental in keeping us on the right track and in critiquing our efforts. The report has been improved greatly by their criticism. David Fogel, my colleague at UICC, and Douglas Thomson, a coauthor of the UDIS report and presently doing research at UICC on probation services, reviewed the final report and made a number of helpful suggestions, as did two

outside reviewers, Jacqueline Cohen of Carnegie-Mellon University and Robert Figlio of the University of Pennsylvania, whose careful critiques led to a number of changes in the report.

A special word of thanks is in order for Lawrence Greenfeld, our NILECJ project monitor. Maintaining the delicate balance between dual roles as advisor on substantive issues and monitor on administrative issues is not an easy task, but is one that Larry accomplished with tact. In addition, he kept us informed of other NILECJ projects that related to ours; he assisted us administratively when the occasion warranted it; and his advice on our research was of great benefit to the project.

Finally, our sources of information. We visited correctional officials in eleven states and were received coridally by all. They provided us with valuable insights into their operations, which helped us to attune our findings to their needs. Our search for data to test the methods also met only with cooperation from our colleagues in various agencies throughout the country. In particular, we wish to thank George Cox, Director of Research of the Georgia Department of Offender Rehabilitation; Peter Hoffman and Barbara Stone-Meierhoefer, Director of Research and Research Analyst, respectively, of the U.S. Parole Commission; Howard Kitchener, Director of Research of the U.S. Bureau of Prisons, and Annesley Schmidt, a former staff member now with NILECJ; Cheryl Ruby and Paul Litsky, Research Associates on the Uniform Parole Reporting Program at the National Council on Crime and

Delinquency Research Center; and Ann Witte, Associate Professor of Economics at the University of North Carolina at Chapel Hill. We hope that we have furnished them with new ways of looking at correctional effectiveness. At least, this has been our goal; the measure of effectiveness of this report will be in terms of the extent to which our recommendations and methods are adopted.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Recent evaluations of correctional programs have been the subject of considerable controversy. Such controversy is not new -- theories of punishment go to the root of the relationship between man and state, and have been critically examined from the time of Plato to the present day. But in the past the debate has normally been confined to academic and professional circles; the present controversy has been placed squarely in front of the American public. Correctional researchers have testified before Congress, have been quoted in the press and in news magazines, and have been interviewed on national television concerning their findings. While evaluations of correctional programs are normally not grist for the media's mills, the findings have been unsettling enough to cause people to question some of the basic goals and assumptions of corrections and correctional policy.

Two evaluative efforts in particular have sparked this interest. One, actually a survey of correctional program evaluations conducted over the past few decades (Lipton, Martinson & Wilks, 1975), has led people to conclude that "nothing works" (Martinson, 1974). The other, an evaluation of juvenile delinquency programs for chronic delinquents (Murray, Thomson & Israel, 1978; Murray & Cox, 1979), has led people

to conclude that major reductions in delinquent activity can result from institutionally-based and community-based correctional interventions (and even from probation -- see Empey & Erickson, 1972). In other words, almost any correctional intervention, from probation to incarceration, seems to work with juvenile offenders.

These seemingly inconsistent findings are not inconsistent. The Lipton et al survey found that recidivism continued after offenders were released from programs; Murray et al also found that juvenile offenders continued to recidivate after release, but at a lower rate than before. In both studies, however, the findings did not appear to depend greatly upon the nature of the correctional program. Taken together, these findings have led correctional policymakers to conclude that the rehabilitative model of corrections is suspect.

Many who spurn rehabilitation as a viable goal of corrections have adopted in its place just deserts as the goal (von Hirsch, 1976; Fogel, 1979). In a major review of research on rehabilitation (Sechrest et al, 1979), it is argued that this shift in goals is not based on philosophical differences. Rather, it is based on the fact that no programs that have been evaluated thus far seem to result in any noteworthy success in rehabilitating offenders. This, they feel, may be due more to the types of programs that have been tested than to the failure of rehabilitation in general. They note that most of these programs are institutionally-based (because it is easier to run

programs in prison than outside the walls), which may not be as conducive to rehabilitation as external programs (Sechrest et al, 1979: 95). In short, they believe that the trouble with good research on rehabilitation (as Gilbert K. Chesterton said about Christianity) is that it has never been tried. Their strongest recommendation is that "research on offender rehabilitation should be pursued more vigorously, more systematically, more imaginatively, and more rigorously" (p.10).

Methodology Research

One aspect of the rigor that they recommend for rehabilitation research pertains to the development of new methods of analyzing correctional programs (p. 105). The analytic methods used in evaluating correctional programs and in measuring relevant behavioral phenomena are quite limited. This report focuses on these very issues. It includes both a critique of present methods and the development of new methods for assessing correctional programs.

To do this we must first say what we mean when we conclude that a program is effective, or that one type of offender did better than another. Effectiveness and improvement are measured with respect to goals, in this case the goals of correctional interventions. Chapter 2 is devoted to a discussion of these goals, specifically from the standpoint of evaluating the extent to which they are achieved.

However, we do not intend to examine all of the evaluative measures used to assess the achievement of these goals. Instead, we have concentrated our efforts on an intensive examination of the most commonly used correctional measure, recidivism. But recidivism is not without its drawbacks as an outcome measure. In Chapter 3 we describe the arguments against using recidivism as a measure of correctional effectiveness and discuss instances in which its use is appropriate.

In both of the studies cited previously (Lipton et al, 1975; Murray & Cox, 1979) recidivism was used as one of the primary outcome measures. Chapter 4 describes the way recidivism was used in these evaluations and the reasons we have for doubting the validity of their findings. It demonstrates how methodological considerations can have a major effect on conclusions.

Other factors that affect the conclusions of a correctional evaluation are the characteristics of the jurisdiction in which the study takes place. Variations in the criminal laws, in the organization of parole and correctional agencies, and in policies and procedures from state to state (and within states as well) mitigate against making direct comparisons across jurisdictions. Chapter 5 describes some of the problems that limit the utility of such comparisons.

One of the arguments against using recidivism as a measure of correctional effectiveness is that its definition seems to be so fluid. Rarely do two studies in different states use the same definition of recidivism. In Chapter 6 we describe how the information available to the researcher limits the definition that can be used. We then consider the ethics of using arrest as the indicator of recidivism. The various ways that recidivism has been defined are also described. Chapter 6 concludes with a suggested schema for categorizing recidivism definitions so that they will be more comparable from study to study.

A common definition of recidivism and similarity of organizational characteristics are necessary to compare the results of two studies. But to understand the results one must also consider the assumptions made about the recidivism process that are implicit in the analytic methods used. Chapter 7 explores these assumptions and shows why a specific family of recidivism models is preferred.

These models are used to estimate parameters of the recidivism process. The estimation procedure, described in Chapter 8, is then used to analyze a number of different correctional programs. Data obtained from Illinois, North Carolina, the U.S. Bureau of Prisons, the U.S. Parole Commission and the Uniform Parole Reports Program produce results which demonstrate the applicability of the methods we propose. They are reported on in Chapter 9.

The analytic methods do have a number of limitations, some of which are made evident in the applications described in Chapter 9. These limitations are discussed in Chapter 10, and a description of our continuing research in this area is given.

CHAPTER 2

CORRECTIONAL GOALS AND THEIR EVALUATION

Correction or rehabilitation of the offender is but one of the goals which society places on prisoner custodial and treatment programs. Sechrest et al (1979: 18) list seven goals of criminal sanctions:

"1. to deter the offender from offending again by punishment or fear of punishment (without necessarily changing him or her in any other way);

2. to deter others from behaving as the offender has;

3. to incapacitate the offender and thus deprive him or her of the opportunity to offend again for a given period of time;

4. to forestall personal vengeance by those hurt by the offender;

5. To exact retribution from the offender and so set right the scales of moral justice;

6. To educate people morally or socially;

7. To rehabilitate or reform the offender."

These goals are described briefly in this chapter. They are value-laden goals: they point towards producing beneficial effects on society or on the future conduct of offenders. Still other correctional goals are more in the nature of administrative or managerial

objectives. They include sentence equity; safety and humaneness in institutional care; and protection of the incarcerated, the institution staff and the public from incarcerated offenders. These goals are also discussed in this chapter.

Our discussion of correctional goals is not meant to be a full exploration of correctional and penal philosophy. This is the province of legal and moral philosophers, and is well beyond the scope of this report. But evaluation of correctional outcomes with respect to these goals is within the scope of this report; and it is from an examination of these goals that we should develop outcome measures for evaluative purposes.

Rehabilitation

The very word "rehabilitation" contains certain implications about correctional programs. It implies that:

- incarcerated individuals have problems, problems which are a direct cause of their criminal behavior;

- correctional program personnel can diagnose these problems accurately, and have apppropriate treatments available for the individuals;

- the treatments will be properly and appropriately applied;

- the problems will be "corrected" (or at least mitigated) as a result of the treatment; and

- the individuals' criminal behavior will begin to diminish as a result of mitigating the problems.

This sequence of implications is what is normally meant when "rehabilitation" or "correction" is considered as the goal of the criminal sanction. Such a logical construct is highly appropriate for medical treatments, where the causal links are physical and physiological and can be traced with appropriate instrumentation. But it is a heavy burden to place on correctional treatments. Too little is known about behavioral change or the causes of criminal behavior to make this sequence of implications apply to corrections. Yet it may still hold true for some programs and some offender types. Evaluations focused on the rehabilitative goal gauge the extent to which this logical construct, this model of the correctional process, actually holds.

But despite the fact that most offenders are sent to "correctional" institutions or participate in "correctional" programs, and despite the fact that the agencies administering these institutions and programs are called departments of corrections or offender rehabilitation, this title is more hopeful than descriptive. One could hardly expect otherwise: programs and agencies should not be held

responsible for what is basically an individual's own decision to continue to break the law, especially when such behavior occurs in environments beyond the control of correctional organizations.

An evaluation based on the goal of rehabilitation would be primarily a process evaluation;[1] that is, it would be concerned primarily with validating the logical construct described above: Did the offender have problems? Were they a direct cause of the criminal behavior? And so on. To determine the nature of these problems, psychological testing of the offenders would be a major component of the evaluation, both before and after treatment. In addition, an assessment of the diagnostic and treatment capabilities of the correctional agency would have to be undertaken, to determine whether the correctional program can actually be implemented in the agency.

Performing an evaluation of this sort is not a simple task. Psychological tests are not entirely reliable, and would be even less so when applied in a a prison setting. Consequently, evaluations focusing on corrections or rehabilitation often deemphasize the process evaluation to concentrate on the impact evaluation:[2] measuring the extent of problem amelioration in the treatment group, and comparing it to the improvement noted in a similar control group. If there is a difference in improvement it may be attributable to the treatment, regardless of the nature of the process or the way the treatment worked. This evaluation approach, of course, is considerably

less satisfactory than one incorporating a process evaluation: it is important to know how a program worked (or why it didn't) to determine its applicability under different circumstances and for different populations.

But even this type of impact evaluation is not always used. One often finds rehabilitation programs being evaluated not on the basis of what the treatment is expected to accomplish, but rather on the basis of recidivism. The question thus addressed is not, "Did the treatment mitigate the problem?", but, "Did the program (somehow) reduce the post-release criminality of its participants?" Recidivism, then, is employed as a measure of effectiveness for the goal of rehabilitation.

As von Hirsch (1976: 11) points out, "In the literature of rehabilitation, there is often considerable ambiguity whether the aim is to reduce recidivism (a form of crime prevention) or to help the offender with his own problems (a paternalistic goal). But treatment programs have generally been tested by measuring their effects on recidivism -- suggesting that the goal of reducing recidivism is actually the primary one."

This latter approach is the one recommended by Sechrest et al (1979). Not only do they point out that recidivism is the "traditional measure" for evaluating rehabilitation programs, but they state that it is "the sole criterion against which rehabilitation ultimately

must be measured" (p.21; emphasis added). This point of view is troublesome. Looking only at the bottom line, recidivism, without consideration of how the program effected this outcome would appear to be a shortsighted approach to program evaluation. We do not dispute the importance of considering recidivism in evaluating rehabilitation programs (after all, in this report we develop new methods of doing so); however, we question the extreme position they have taken. Other evaluative measures are often of much greater benefit in understanding a program's effectiveness than is recidivism alone.

The Goal of Deterrence

Two different types of deterrence were distinguished by Zimring and Hawkins (1973), general deterrence and special deterrence. General deterrence is the general reduction in criminal activity attributable to a planned intervention. The intervention may be one based on criminal justice system actions -- e.g., more police or increased penalties (Press, 1971; Campbell & Ross, 1968) -- or it may be based on individual actions -- e.g., Operation Whistlestop (Reed, 1978) or community organization (Cirel et al, 1977). Special (or individual) deterrence is the reduction in criminal activity by a specific offender or group of offenders, as a direct consequence of a penalty, incarceration or some other sanction. Deterrence implies that potential offenders have somehow been convinced that initiating

(or continuing) criminal behavior is not in their best interest, either because of the penalties imposed on others (general deterrence) or on themselves (special deterrence) after committing crimes. A proper evaluation should thus be based upon the extent to which a reduction in criminal activity resulted from a specific action.

There are a number of thorny theoretical and methodological problems involved in estimating the general deterrent effect of different policies. For example, Ehrlich's (1973, 1975) conclusions about the general deterrent effect of capital punishment on homicide have been hotly debated (Baldus & Cole, 1975; Bowers & Pierce, 1975; Forst, 1976). The relative rarity of both homicides and executions makes the analysis quite sensitive to very small changes in the number of events. Furthermore, both the increase in crime in general -- as well as in the homicide rate -- and the decrease in the number of executions over the past few decades might well have been due to factors unrelated to each other: correlation between the two time series does not imply causation. In addition, changes in the assumptions implicit in the analytic methods used make a considerable difference in the conclusions; for a review and critique of these issues, see Blumstein, Cohen and Nagin (1978). Evaluation of the general deterrent effectiveness of an intervention is thus not a simple task.

Evaluation of special deterrence is considerably more straightforward, since one only need trace the future criminal careers of a

specific group of offenders. The recidivism characteristics of this group, then, is the evaluative measure for determining the extent to which special deterrence has been achieved. However, since this same measure is used for rehabilitation one cannot always be certain which of these goals is being achieved: it may be that both the carrot (rehabilitation) and the stick (special deterrence) are at work.

The Goal of Incapacitation

Incapacitation refers to the fact that an incarcerated offender cannot commit a crime -- at least against the general public -- while he is incarcerated. To evaluate the extent to which crime is reduced as a consequence of incapacitation, one must estimate the number of crimes that would have been committed had the offender been incarcerated for a shorter period of time (or not at all). One can then show how different sentencing practices affect the total crime rate. Understanding what would have happened had the offenders been freed is a crucial component of this procedure; in recent years a number of models of criminal activity have been proposed; see, for example, Avi-Itzhak and Shinnar (1973), Shinnar and Shinnar (1975), and Greenberg (1975). Their claims as to the incapacitative effect of prison on the crime rate have been shown to be based on a variety of simplifying assumptions (Cohen, 1978), assumptions which may make the mathematics more tractable but may not typify offender behavior. For

example, to simplify the equations it is assumed that offenders commit crimes at a (statistically) constant rate over time, and that they all have the same criminal career length. Neither of these assumptions is based on empirical evidence, and the sensitivity of the findings to these assumptions has not been explored.

More recently, empirical investigations into patterns of criminal activity have been conducted (Petersilia et al, 1977; Chaiken, 1977; Greenwood, 1979). Models based on these patterns should make possible an improvement in estimating the incapacitative effect of incarceration on the crime rate.

One of the unknown factors in such modeling efforts is the extent to which post-release criminality is affected by incarceration. That is, will an individual upon release try to "make up for lost time" by increasing his criminal activity, in effect negating the incapacitative effect? Again, an investigation of post-release criminality would be informative, so recidivism is useful in studying certain aspects of the incapacitative effect of incarceration.

The three goals discussed above are concerned with the effect of the criminal sanction on mitigating problems that caused criminality, or otherwise reducing their criminal behavior during or after incarcertaion. The next three goals we discuss -- forestalling vengeance, providing moral and social education, exacting retribution -- are concerned more with the effect of the criminal sanction on the non-

criminal segment of society: keeping victims from demanding an eye for an eye, reinforcing the norms and ethics of our society, demonstrating that those guilty of harming others do receive punishment. Evaluating these goals would necessarily involve population surveys; since they are not program-specific, the evaluation of these goals will not be discussed herein extensively.

Forestalling Vengeance

We entrust the criminal justice system with a monopoly on the legal use of coercive force. In return we expect that we will be protected from crime, and that the criminal justice system will punish offenders in the event that the protection proves ineffective and the offenders are known.

Since most offenders cannot be identified by the victims, personal vengeance cannot be carried out for the most part. However, this does not mean that all vengeance is forestalled; it may be practiced vicariously. Many television programs use this theme; and the popularity of the recent movie, "Death Wish" -- in which a man whose wife and daughter were assaulted goes after muggers and rapists -- suggests that criminal sanctions do not completely eliminate all feelings of personal vengeance. In addition, vigilante groups have sprung up in many communities to augment what they see as ineffectual police activity (Marx & Archer, 1971, 1973); and in Brazil, where the crime

rate has grown dramatically in recent years, lynchings of criminal suspects are not uncommon (Hoge, 1979).

Evaluating the effectiveness of criminal sanctions in forestalling vengeance is not a trivial task. In fact, it may not be possible to assess. Public confidence in the criminal justice system is one possible measure, but it is a weak indicator of the extent to which private vengeance is eschewed.

Moral and Social Education

The criminal law and its accompanying punishment serve notice to the public exactly what behaviors are proscribed.[3] This announcement is an integral aspect of general deterrence: potential offenders cannot be deterred from committing crimes if they are unaware that the behavior is defined as criminal. In that sense, the educational goal of the criminal sanction can be seen as an intermediate step between the sanction and deterrence. For crimes considered *mala in se*, such as murder, rape and robbery, there is rarely any question about this linkage; but for *mala prohibita* (gambling, prostitution, illegal drug use) the educational goal does not necessarily lead to deterrence. The primary distinction is perhaps between the *moral* and *social* aspects; insofar as there is dissensus concerning the imposition of another's moral position on oneself the educational goal of the criminal sanction is undermined.

Retribution and Desert

Criminal sanctions are often used, not just as a means of protecting the public by reducing the crime rate (as in correction, deterrence and incapacitation), but to punish the offender because of the intentional harm that he visited upon the victim. Retribution and desert focus on this concern. von Hirsch (1976: 46) discusses the perjorative connotation of retribution, which seems to imply revenge, vindictiveness and punishment out of proportion to the offense; while desert implies a measured punishment meted out rationally.

It is difficult to develop a means of evaluating the extent to which these goals are achieved. As Morris (1974: 75) points out, "Desert is, of course, not precisely quantifiable." It might be possible to gauge the public's perception of degree of harm caused by each offense, and to compare their perceptions with punishments for those offense. Perhaps the study by Sellin and Wolfgang (1964) on the public's perception of offense seriousness comes closest to the mark, but even in that work it is impossible to disentangle retributive feelings from assessments of seriousness based on the need for deterrence or rehabilitation.

Proportionality and Sentence Equity

Retribution and desert are related to the concepts of propor-

tionality -- letting the punishment fit the crime and the past behavior of the offender -- and sentence equity -- giving like punishments to like offenders who commit like crimes. An evaluation of the achievement of these goals would be based on the schedule of sentences as served (in contrast to the sentences handed down by the court). Thus, an evaluation would involve looking at the entire criminal justice system: legislation (what sentences are prescribed, and why); police discretion (of those found committing an offense, who gets arrested); prosecutorial discretion (of those arrested, who is charged, indicted and tried for the offense, who for a lesser offense, whose case is nolle'd); sentencing discretion (who is sentenced to prison, who is given probation, who is sentenced to "community service"); and correctional discretion (who does easy time, who gets paroled early).

Recent works that have been concerned with these goals include studies by the American Friends Service Committee (1971), Fogel (1979) and von Hirsch (1976). Their influence has been felt in the enactment of new laws and criminal codes in a number of states, in which the discretion of judges and parole authorities in changing sentences is curtailed. Efforts are currently under way to evaluate the impact of these laws.

Institutionally Focused Goals

Other goals focus on the administration of custodial institu-

tions. They include the provision of adequate food, housing and health care, safety of the residents, the prevention of escapes. These goals can also be evaluated, based on existing standards in each of these areas, and on the extent to which the standards are met. The Law Enforcement Assistance Administration has funded a Correctional Standards and Accreditation Project toward this end (Corrections Digest, 1979).

- - - - -

As we have seen, the multiplicity of goals that are the putative responsibility of the correctional system require a number of different procedures to evaluate the extent to which they are reached. Different types of data need to be collected and different analytic techniques need to be used for their evaluation. However, consideration of the post-release behavior of offenders is useful in assessing the extent to which two goals of corrections have been achieved: rehabilitation and special deterrence. Our research has concentrated on this measure of correctional effectiveness, on the development of techniques to evaluate the post-release behavior of offenders.

NOTES

1. Every program has (or should have) a rationale, a raison d'etre, based on an empirical or theoretical body of knowledge, that leads one to believe that the program will have a certain impact. A process evaluation is an assessment of the extent to which the rationale holds. It is often difficult to explain the logical connection between the resource input and the outcome, especially if the program is based on implicit assumptions: "gut feelings" are not easily translated into cold logic.

2. An impact evaluation is an assessment of the extent to which the program goals are achieved, regardless of whether they were achieved by the program as described in the rationale, by the program in some unforeseen way, or by some other means. Use of a control group will tell the evaluator how much impact the program had on the population under study; but a process evaluation is necessary to determine how the program caused the impact.

3. A recent example of the deterrent effect of publicity can be found in studies of the Massachusetts gun control law, passed in 1975. The law, no more restrictive than laws in other states, resulted in a significant drop in certain gun-related crimes (Deutsch & Alt,

1977; Pierce & Bowers, 1979) in Boston. The major difference between the way this law was enacted was that it was accompanied by media publicity (which has continued). Potential offenders are regularly warned of the consequences of illegally carrying concealed weapons. It is also very likely that this campaign's effect is not restricted to potential offenders: the main benefit may be in its encouragement of police, prosecutors and judges to enforce the law fully.

CHAPTER 3

RECIDIVISM AS A MEASURE OF CORRECTIONAL EFFECTIVENESS

Recidivism, or post-release criminality, is the outcome measure used most frequently in evaluating correctional programs. In fact, some consider recidivism to be "the sole criterion against which rehabilitation ultimately must be measured" (Sechrest et al, 1979:21). But it has often been used indiscriminately without regard for its appropriateness or limitations. In this chapter we discuss some of its more salient limitations. First, we explore the values that are implicit whenever recidivism is used as an evaluation criterion. We next delve into the more practical problems of using recidivism as a measure of correctional effectiveness. Alternative evaluation criteria are then discussed. Finally, we discuss what recidivism can tell us and when the use of recidivism is appropriate as an outcome measure.

The Values in Evaluation

Correctional evaluations often make the implicit assumption that it is the individual offender who needs correcting; yet some will argue that the social and political system that created the criminogenic environment should be corrected. The increase in crime would thus be seen as a measure of the extent to which society is failing the indi-

vidual. According to this argument, recidivism should be looked upon as an indicator of deficiencies in society in general (and in the criminal justice system in particular), not only of deficiencies in the individual offenders.

The assumption that the offender needs correction may be incorrect for other reasons. No doubt there are some offenders whose criminal activity is promoted in some way by correctible defects: inability to read, lack of employable skills, personal or family problems. But some may choose to commit crimes because it is easier than working a straight 9-to-5 job; some because they enjoy risk-taking; some because of peer pressure. These reasons for committing crimes may be considered defects by some, but not necessarily by the offenders.

A concentration on recidivism conveys the implication that we can do something to reduce the post-release criminality of offenders, that we just haven't found the right combination of treatment modalities yet. In other words, the offender is not responsible for his subsequent acts; it is we who have failed to provide for his rehabilitation. This, of course, presumes that it is easy to change a person's behavior. But as Wilkins (1969: 7) points out, "If it were possible to change from 'bad' to 'good' without much effort, changes from 'good' to 'bad' could probably be effected as simply and perhaps would involve a larger proportion of the population."

To label a program a success or failure on the basis of its participants' future criminality is to distort the true value of many programs. Good programs may be curtailed or eliminated because recidivism is the dominant measure of correctional effectiveness. If an illiterate offender has learned to read while in prison, but committed a crime after release because he had a heroin habit or because he was unemployed, this crime has nothing to do with the program's effectiveness. The program may have been quite successful, but success at one level does not automatically (or immediately) lead to success at another level. The underlying assumption that alleviating an individual's problems will cause him to turn away from criminal activity unfortunately does not always hold.

But it often does hold; the criminal activity of some offenders may be reduced because of the program. However, one should not expect an instant and total conversion on the part of these offenders. Habits of a lifetime should not be expected to disappear immediately, no matter how successful the program. Since recidivism measures are normally based on post-release behavior immediately after release, the program's beneficial effects may be underestimated.[1]

When a program is evaluated using recidivism, i.e., failure, rather than using criteria that highlight success, a particular set of values may be fostered among program evaluators. Their attention is given to program participants only when they fail; they do not follow

up and report on the participants' successes. This can color a program's evaluation -- program evaluators who are asked to document failures rather than successes may approach their task from a pessimistic viewpoint.

Critics of correctional evaluations have suggested looking beyond the goals of corrections to the goals of the criminal justice system in general, and evaluating programs on the basis of their contribution to these goals. One might consider a goal of the criminal justice system to be the reinforcement of societal values; another goal might be the reduction of harm to society due to crime. Not addressing these goals and focusing on recidivism is thus seen as a shortsighted policy because its relevance to these overall goals may be marginal.

Practical Difficulties in Recidivism-Based Evaluation

A number of practical difficulties arise in the evaluation of correctional programs. Some are present regardless of the outcome measure used; for example, one can rarely run a controlled experiment in corrections. Other difficulties relate to the choice of recidivism as the outcome measure: lack of a standard definition and poor data quality are common problems in studies of recidivism. These and other problems are discussed below.

Experimental Design. The assumption that a particular correctional program causes the behavior of its participants to change is

often doubtful. Even when all threats to validity have been accounted for, as in a well-designed correctional experiment, attribution may be faulty. For example, success may be due more to the personalities of the staff running the program than to the nature of the program: given that same staff, any program would show signs of success. One might say that this is true success anyway, since the staff are part of the program. But one normally runs correctional experiments to test correctional treatments for widespread use. If they are successful only when implemented by a dedicated staff, the "external validity" of the experiment (i.e., its generalizability to other settings) is suspect.

But true experiments are the exception in correctional environments. Sechrest et al (1979: 60) argue the case for true experiments, but an innovation of this nature can encounter strong resistance. It is rarely possible to form equivalent experimental and control cohorts when doing correctional research. Prisoners are protected from being coerced into "volunteering" for experimental programs where they are subject to manipulation by researchers.[2] Therefore, quasi-experimental designs[3] are the norm in correctional research; and quasi-experiments contain many threats to validity.

For example, in a quasi-experiment one may try to match the cohort volunteering for an experiment with a cohort of those who do not volunteer. Even if both cohorts have comparable distributions of age,

race, prior offense records, education, etc., there is one variable that cannot be accounted for -- voluntarism. And prisoners volunteer to participate in experimental programs for a variety of reasons -- boredom, a desire to impress the parole board, a real interest in the program, a desire to better themselves -- some of which are quite relevant to correctional success. Blumstein and Cohen (1979) address this issue to some extent in a quasi-experimental evaluation of a prison-based college educational program. Thus, a difference in recidivism rates between cohorts in a quasi-experimental program evaluation may be attributed to the program but is actually due to the hidden difference in cohorts. "If the matching techniques fail to account for all initial differences that have a bearing on treatment outcome, a study's findings will be systematically biased" (Rezmovic, 1979: 167).

Recidivism definitions. There is no consistent definition of recidivism. One program may use a one-year follow-up, another six months.[4] Follow-up time may be computed starting with release from prison, or with release from parole. The recidivating event may be a technical violation of the conditions of parole, or it may be return to prison. There are so many possible variations in the method of computing recidivism that one doubts if more than a handful of the hundreds of correctional evaluations are truly comparable. Nor is there any way of deciding which of the many variations is most ap-

propriate for a given set of circumstances (Sechrest et al, 1979: 73).

Data quality. Recidivism data are based on reported events, not all events. Studies of crime reporting behavior (LEAA, 1977: 9) have indicated that only about half of the crimes reported to Census Bureau interviewers are reported to the police. When studying crime, then, one can supplement data on crimes reported to the police with a survey of the victims of crime to get a more complete picture of the crime problem. However, we cannot expect a similar source of supplementary information for recidivism. The victims of crime normally do not know who committed the crimes, so a victimization survey cannot be used to "tag" a specific ex-offender with a specific post-release crime -- unless, of course, the ex-offender is rearrested. And a survey of ex-offenders would probably not be helpful; it is unlikely that offenders will respond to a survey questionnaire to document crimes they got away with.

There is an additional problem in using only officially reported events in computing recidivism statistics. The data are very sensitive to policy shifts within the data-collecting agencies. For example, parole officers may be told to be lenient with their parolees (if the prisons are overcrowded) or to tighten up on parole (if the parole agency has been receiving unfavorable publicity). The recidivism rates in the two situations would probably be quite diffferent.

Changing police arrest policies may have a similar effect. The extent to which policy shifts of this kind affect recidivism statistics is not known.

Because of the measure's variablity (described in greater detail in Chapters 5 and 6) one cannot state with any degree of assurance whether a given recidivism rate is high or low: there is no "standard" recidivism rate as there is a normal body temperature. Therefore, recidivism can only be used as a comparative measure. Analogizing recidivism to body temperature is apt for another reason. Both are crude measures of a phenomenon, but neither can be used to diagnose the subject of the investigation -- additional variables must be measured for diagnostic purposes.

Another problem with recidivism data is that recidivism events are all given the same weight. Everything is either black or white, success or failure; there are no shadings of gray, no middle ground: "A great deal of information is lost when something as complex as possible criminal activity that may or may not culminate in detection, arrest, and conviction is finally expressed as a simple dichotomy" (Sechrest et al, 1979: 71).

One approach to this problem is to weight the event according to its perceived "seriousness" (Moberg and Ericson, 1972), but this approach has deficiencies, too. An individual may have his parole revoked ostensibly for a technical violation, but actually because he

committed another crime. Or he may be arrested for a crime he did not commit, is subsequently released, but is considered a recidivist because an arrest is one of the events used to define recidivism. Considering this lack of consistency, one wonders how useful it would be to use a measure more complicated than a simple dichotomy.

Other Measures of Effectiveness

Recidivism may be thought of as a measure of success as well as failure. That is, those who do not fail could be considered successes. But this is a very limited and pessimistic view of success -- not (having been found) getting into trouble. A measure of success should be based on positive accomplishments, not on the absence of negative findings.

However, it is failure that is recorded by agencies, not success. Evaluators are inclined to use the data collected by these agencies for their evaluation because it is much easier than collecting new evaluative data. Despite the well-known problems in using official data for research purposes (Kitsuse & Cicourel, 1963; Maltz, 1972), they are used because they are there. Their existence inhibits the collection of more relevant data, since the cost of collecting reliable and valid data is quite high. Aside from financial limitations, very often evaluators are not given access to better sources of data for considerations of privacy and confidentiality.

Measuring success is much more intrusive than measuring failure. To determine how successful an individual is, one would need to investigate his employment situation, his family situation, and other aspects of his personal life. One cannot expect ex-offenders to volunteer to give this information to evaluators, just because the evaluators feel the need to measure success. Even if all ex-offenders could be counted on to provide the information, one would not be able to gauge the extent of dissembling and exaggeration. And the cost of data collection and verification would be quite high if the data had to be collected especially for the evaluation.

But this type of information is routinely collected by parole and probation officers. For special evaluations these officials could use standardized collection instruments and procedures for gathering information on employment, educational attainment, family stability, and other relevant variables. Additional procedures, similar to those used by survey research organizations, could be implemented to check on data reliability and validity. For parolees or probationers, then, measuring success may be feasible if sufficient cooperation is obtained from the parole or probation agency.

Measuring success. Success must be defined with respect to a goal. The goal should be defined with some degree of precision; the sources of data used to measure the extent of goal attainment should be specified; and the method of analyzing the data should be specified, since it will also affect the evaluation.

For example, "employment" is a common goal of correctional programs. One can use a number of different definitions to determine the extent of its attainment: number placed in any job; number placed in a job employing the skills acquired in the program; number placed in a job who remain for a specified follow-up time; number placed, controlling for employability and/or for the local unemployment rate; etc. The specific measures used should reflect the perceived goals of the correctional program.

Similarly, one can use a number of different sources of data to measure the extent of goal attainment. Among them would be reports of telephone conversations with program participants; telephone contacts with the participants' employers; and visits to the workplace to assess job placement firsthand. One would expect different degrees of reliability and validity and different costs of collecting data for each of these data sources.

Employment data can be analyzed in a number of different ways. One can look upon employment as a binary variable (employed/not employed) or a continuous variable (hours worked per week). One can look at an individual's employment history over time, or one can select a single point in time (say, six months) and ascertain whether he is employed then, or whether he has been employed throughout that time, or the average number of hours he worked during that time. Each method of analysis will produce different evaluation findings.

Each operationalization of the goal "employment" has its own strengths and weaknesses. One should try to determine the degree to which the measured quantity actually represents the goal -- for example, do program personnel tell the participants, "Just stick with this job for six months, so we can get a good evaluation."? This consideration is crucial in gauging how well the goal was achieved (impact evaluation) and how effective the program was in contributing to its achievement (process evaluation).

A complete study of correctional program effectiveness would explore the way correctional goals are operationalized. It would compare the strengths and weaknesses of each operationalization, and would describe the types of program for which each appeared applicable. For example, one would use different definitions of "employment" to evaluate an auto mechanic training program and to evaluate a program teaching basic job skills ("get to work on time, wash daily, dress neatly"). Operationalization, therefore, is based on the nature of the correctional program [see Grizzle (1979) for a discussion of performance measures for evaluating correctional programs].

- - - - -

We see, then that there are many valid criticisms of recidivism as a measure of effectiveness. It is difficult to measure with precision; even if it could be measured precisely, there are no set standards with which to compare measurements; and even if standards existed, the interpretation of the findings might very likely be flawed. But problems of this nature are common to virtually all social phenomena. The measurement of poverty, educational attainment, employment, self-esteem, socioeconomic status, social structure, peer group relationships, etc., is no less difficult than the measurement of recidivism. Doubtless none of these measures will ever be as precisely defined as physical measures such as temperature or pressure. One is therefore left with the option of making do with imperfect measures or of not measuring social phenomena at all. This is not to say that all uses of recidivism are appropriate, especially in evaluating rehabilitation programs.

But recidivism is still quite useful as a program measure, as a correlate measure of secondary importance if not always as a primary measure of effectiveness.[5] When used in this way recidivism can be informative. In conjunction with other measures of effectiveness it can be used to evaluate programs where rehabilitation is an achievable goal, as in juvenile delinquency programs. It can be used to evaluate selection criteria for community-based correctional programs. It can be used to estimate the extent of special deterrence. It can be used

in modeling recidivism as a feedback process (Belkin, Blumstein & Glass, 1973). It can be used to analyze certain questions relating to the termination of criminal activity: is there a certain age when most offenders stop their criminal activity? How does it vary by offender characteristics, by type of criminal career path, by other factors? Thus, recidivism can be a useful measure to study the characteristics of offenders.

There also are instances where its use can lead to inappropriate conclusions. In the next chapter we discuss two recent studies of correctional programs that used recidivism as one of the measures of correctional effectiveness.

NOTES

1. It may be that criminality does decline after release from a program. However, extreme care should be taken in pretest-posttest evaluations to ensure that the decline is real and not an artifact of the selection process. See Chapter 4 for a discussion of how this artifact arises.

2. It is often moot as to who is manipulating whom. Courses in research methodology rarely spend time on describing the con artists who try to put one over on the researcher.

3. Cook and Campbell (1979: 6) describe quasi-experiments as "experiments that have treatments, outcome measures, and and experimental units, but do not use random assignment to create the comparisons from which treatment-caused change is inferred. Instead, the comparisons depend on nonequivalent groups that differ from each other in many ways other than the presence of a treatment whose effects are being tested. The task confronting persons who try to interpret the results from quasi-experiments is basically one of separating the effects of a treatment from those due to the initial noncomparability betwen the average units in

each treatment group; only the effects of the treatment are of research interest."

4. The National Advisory Commission on Criminal Justice Standards and Goals (1973: 529) recommends a follow-up time of three years. "While this is an arbitrary figure, it is chosen because the few recidivism studies that have followed offenders more than three years have shown that most recidivism occurs within 3 years of release from supervision."

5. Concerning this point we take issue with Sechrest et al (1979: 21), who states, "Criminal behavior, rather than offender growth, insight, or happiness, is the sole criterion against which rehabilitation must ultimately be measured." Offender employability, offender literacy, and even offender growth, insight, or happiness <u>are</u> relevant criteria if the rehabilitative programs were directed toward those ends.

CHAPTER 4

WHAT WORKS?

Recidivism, using one operational definition or another, has been the dominant outcome measure used in evaluating correctional programs. [Chapter 5 describes the recidivism definitions that have been used in studies of correctional programs.] Two of these studies have recently received a great deal of attention. Both came to conclusions which question the efficacy of correctional programs in reducing post-release criminality. One conclusion, based on comparisons of control groups with groups in experimental programs, is that "nothing works", that is, no program seems to be the key to reducing recidivism. The second conclusion, based on comparisons of the pre-intervention and post-release behavior of juvenile delinquents, is that "anything works"; that is, one can almost guarantee a substantial reduction in delinquent behavior with any intervention, whether probation or incarceration, whether the program is rehabilitative or punitive in nature.

Both of these conclusions are quite strong and unequivocal. But their validity is open to question. In this chapter we discuss the background leading to these conclusions and the questions that have arisen about their validity.

"Nothing Works"

Over the past few decades literally hundreds of correctional programs have been evaluated. In an effort to use the information generated by these studies to plan correctional programs, the New York State Governor's Special Committee on Criminal Offenders funded a reanalysis of these studies. This reanalysis was completed in 1971 and published in 1975 (Lipton, Martinson & Wilks, 1975; hereinafter referred to as LMW), but the prior publication of its major findings (Martinson, 1974) had a strong impact in correctional circles. Martinson's "bald summary" of the findings was: "With few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism" (Martinson, 1974; emphasis in the original). This conclusion was based on an evaluation of studies published between 1945 and 1967 that were in the open literature or available from "agencies and individuals conducting evaluation research on treatment in the United States, Canada, and Western Europe" (LMW, 5). Over one thousand studies were identified, but only 231 of the studies met their minimum standards of scientific rigor that would permit them to be reanalyzed. There were 286 separable findings from these studies, classified in Table 4-1 according to the nature of the treatment (eleven types) and the nature of the outcome measure used (seven types). As can be seen, almost half of the findings used recidivism as the outcome measure, and this measure was given the most attention in the book.

Table 4-1 Treatment Methods by Outcome Measures

Outcome Measures (Dependent Variables)	Treatment Methods (Independent Variables)											
	Probation	Imprisonment	Parole	Casework and Individual Counsel	Skill Development	Individual Psycho-therapy	Group Methods	Milieu Therapy	Partial Physical Custody	Medical Methods	Leisure-Time Activities	Total
Recidivism	18	19	18	7	15	12	19	20	4	5	1	138
Institutional Adjustment	0	2	0	1	3	4	6	5	0	2	1	24
Vocational Adjustment	1	0	0	2	5	3	2	0*	0	1	0	14
Educational Achievement	1	0	0	0	9	1	1	0	0	0	0	12
Drug and Alcohol Readdiction	0	0	3	3	1	1	2	1	1	4	0	16
Personality and Attitude Change	3	10	4	3	3	5	21	8	0	9	0	66
Community Adjustment	0	0	0	2	4	1	3	4	1	1	0	16
Total	23	31	25	18	40	27	54	38	6	22	2	286

*This is a unique entry since feasibility or demonstration studies have not ordinarily been included in the survey.

Source: Lipton et al (1975: 8)

This finding sparked considerable controversy (e.g., Klockars, 1975; Martinson, 1975; Martinson et al, 1976). Palmer (1978) provided one of the more comprehensive responses to this and other criticisms of correctional intervention. He pointed out that almost half of the studies cited in LMW had positive or partly positive findings. The fact that no single treatment was shown to work across the board -- for all types of offenders, under all conditions -- should not be taken as a negative finding if it was effective for some types. Palmer contended that it is the search for rapid and glamorous solutions which has caused rehabilitation to fall into disrepute, that what is needed is patience and precision in formulating a correctional research program.

Martinson's criticism extended beyond this interpretation of past research. He took correctional researchers to task for making overly optimistic claims about treatments, claims that were not borne out in subsequent evaluations. And the fact that all of this research had little or no effect on crime rates was construed by him to demonstrate the failure of correctional intervention. Both of these criticisms were seen by Palmer to be beside the point: the fact that claims are exaggerated does not negate a treatment's actual benefits;[1] and the participation of a small group of specific types of offenders in a correctional treatment program should not be expected to reduce criminality among all types of offenders in society.

The findings of LMW were given strong support in two more recent works. Greenberg (1977) surveyed studies published from the late 1960s through 1975, and reached the same conclusion as LMW regarding the effectiveness of correctional programs in reducing recidivism. And Fienberg and Grambsch (1979) were commissioned by the National Academy of Sciences Panel on Research on Rehabilitative Techniques (Sechrest et al, 1975) to reanalyze a random sample of the studies cited by LMW. They also found no cause to doubt the general thrust of the LMW findings, which they characterized as a "reasonably accurate portrayal".

These more recent reviews went further than LMW and looked more critically at the programs being evaluated. In commenting on the nature of correctional studies, Sechrest et al (1979) noted that most of the programs that have been evaluated have been designed to be carried out within institutions; they suspected that rehabilitation research has often been "dictated more by practicalities than by logic" (p.95). Greenberg (1977: 141) found that the descriptions of many of the treatments studied were vague or nonexistent and their theoretical underpinnings were often not made explicit; when they were made explicit, "they tend to border on the preposterous". Whether the treatment was sufficient in intensity or duration was also questioned (Sechrest et al, 1979: 40), nor was it always certain that the integrity of the treatment was maintained. The studies of correctional

treatments were far from free of flaws in design; according to Sechrest et al (1979:60), "The thousands of extant studies on rehabilitation scarcely add up to a single trustworthy conclusion."

What emerges from this controversy is almost as pessimistic as the conclusion that "nothing works." It is that much of the research on rehabilitation completed thus far has been too weak to permit any conclusions to be made. When promising treatments have been found little effort has been made to follow up on them or to attempt to replicate them in other jurisdictions. We lack a coherent body of knowledge about correctional program effectiveness.

"Anything Works"

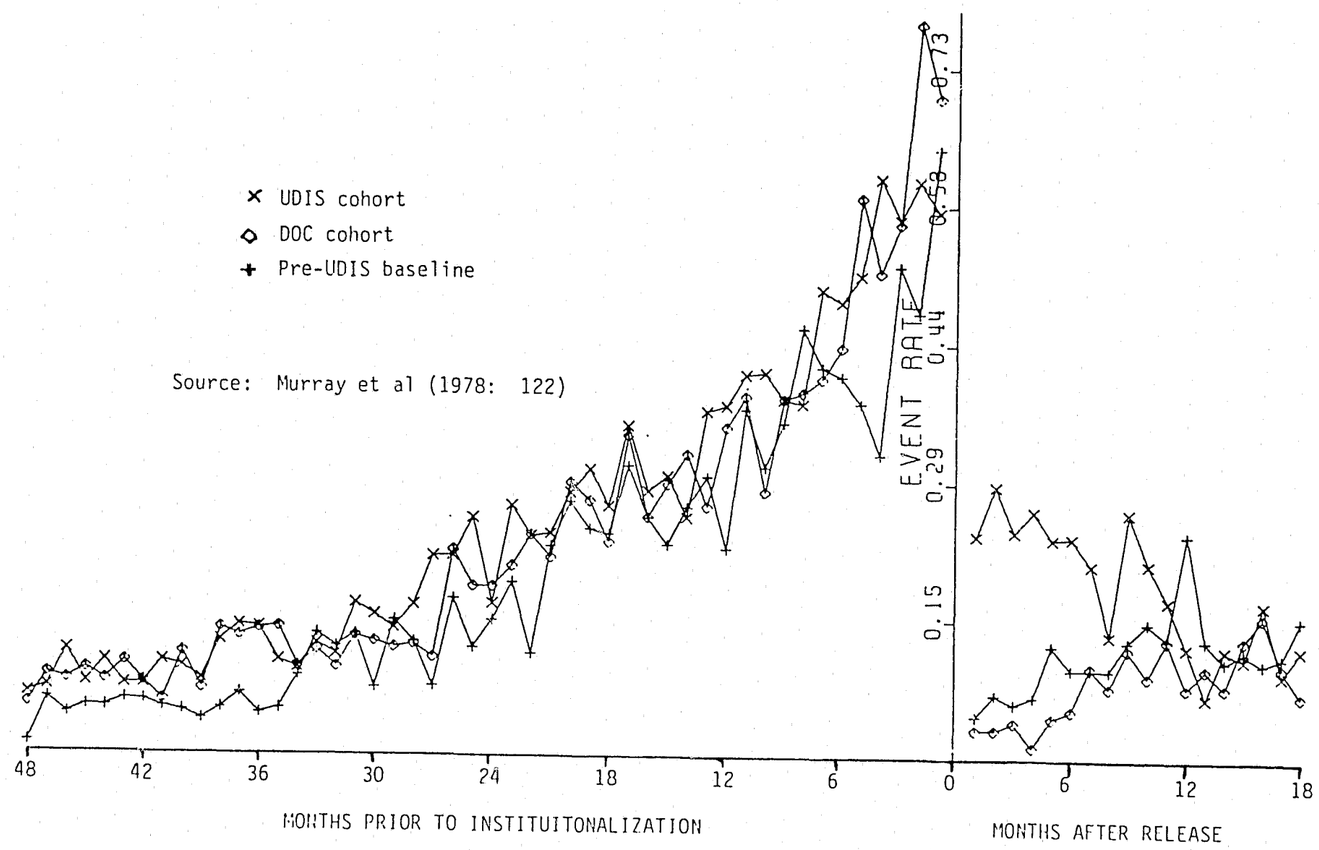
A recent evaluation of a correctional program (Murray, Thomson and Israel, 1978; Murray & Cox, 1979a) generated a great deal of controversy in correctional circles (Murray, 1978a; New York Times, 1978; Gordon et al, 1978; Murray, 1978b; Murray, 1979; McCleary et al, 1979; Kiernan, 1979; Maltz et al, 1980; Maltz & Pollock, 1980; Maltz, 1980). The finding, based on an evaluation of the Unified Delinquency Intervention Services (UDIS) Program of the Illinois Department of Corrections, is that the delinquency rate of chronic juvenile delinquents is decreased substantially following intervention (i.e., there is a "suppression effect") whether the intervention was community-based or institutionally based. Both of the intensive programs

("energetic correctional interventions") they examined achieved a sixty to seventy percent reduction in delinquent activity. Figure 4-1, taken from Murray et al, 1978), is a dramatic representation of this suppression effect.

This is not an isolated finding, and is not due to the quasi-experimental nature of the design of the UDIS evaluation. Empey and Lubeck (1971) and Empey and Erickson (1972) found similar reductions in delinquent activity in an experimental setting, for both experimental and control groups, whether the intervention was probation or incarceration. These findings have given strength to the inference (not drawn by the researchers) that "anything works", that any intervention will produce a profound reduction in criminal activity.

This conclusion has been linked in many minds to Martinson's (1974) conclusion. The net impression that remains from these two conclusions is that doing something to an offender is better than doing nothing, and that no one treatment works much better than any other. On the basis of this impression there has been a revision in the thinking of many involved in correctional policy-making and research. A "get-tough" policy, which has often been cited by many as the solution to crime, is now felt to be strongly supported by current research. This is not the case.

Before discussing this finding, we should note that this evaluation did not use the standard type of operational definition of re-



Source: Murray et al (1978: 122)

Figure 4-1 Arrest Histories for Three Juvenile Cohorts

cidivism, i.e., rearrest (or reconviction) within one year (or some other time period). Rather, they compared arrest rates before intervention (treatment) with rates after intervention. The evaluators used this measure instead of the standard one because they felt it was unrealistic to expect a treatment to effect an immediate and total cessation of delinquent activity. They therefore decided to evaluate UDIs by comparing arrest rates before and after intervention. In every comparison they made they fond a reduction in arrest rate, ranging from a low of 47 percent, for wilderness programs, to a high of 82 percent, for intensive care residential programs (Murray & Cox, 1979: 118). They therefore concluded that offensive behavior was suppressed after intervention.

The "suppression effect" of Figure 4-1 is probably only an artifact of the juvenile justice process. The artifact is quite similar to the "regression to the mean" artifacts discussed in Campbell and Ross (1968) and Campbell and Erlebacher (1970). In those papers, however, the artifact was attributed to the quasi-experimental research designs that were used; the selection artifact that gives rise to an illusory "suppression effect" can occur using experimental (Empey and Lubeck, 1971; Empey and Erickson, 1972) or quasi-experimental (Murray et al, 1978) designs. The artifact is caused by the way juveniles are selected for intervention by juvenile authorities. As Terry (1967) has shown, the greater the number of previous

referrals to juvenile authorities, the more likely a juvenile is to have formal procedures invoked in a given case. So the decision to intervene in a youth's career is based in part on his having a higher than average amount of delinquent activity in the recent past.

It can be shown (Maltz and Pollock, 1980) that even if juveniles have a constant delinquency rate, a curve similar to Figure 4-1 will be generated if the decision to intervene is based on the rule: "Intervene after this offense if a juvenile has had K other offenses within the past T months." Figure 4-2 shows curves for K = 3, 4, 5, 6, and 7 offenses when T is uniformly distributed betweeen 0 and 24 months, superimposed on the UDIS data.

As a prelude to explaining how this artifact might be generated, consider the following scenario: A youth commits a delinquent act. After conferring with a police officer, the victim decides not to make a formal complaint. The youth commits another delinquent act. This time the victim complains, but the police officer still handles it informally. Another delinquent act, and the police officer brings the youth into the police station, calls the youth's parents, and warns them that the next time it will be dealt with more formally. After the next delinquent act the youth is referred to the juvenile probation officer, who also warns the youth and his parents. The next delinquent act causes the youth to be brought before the court, but he is put in a diversion program rather than put on trial. It is the

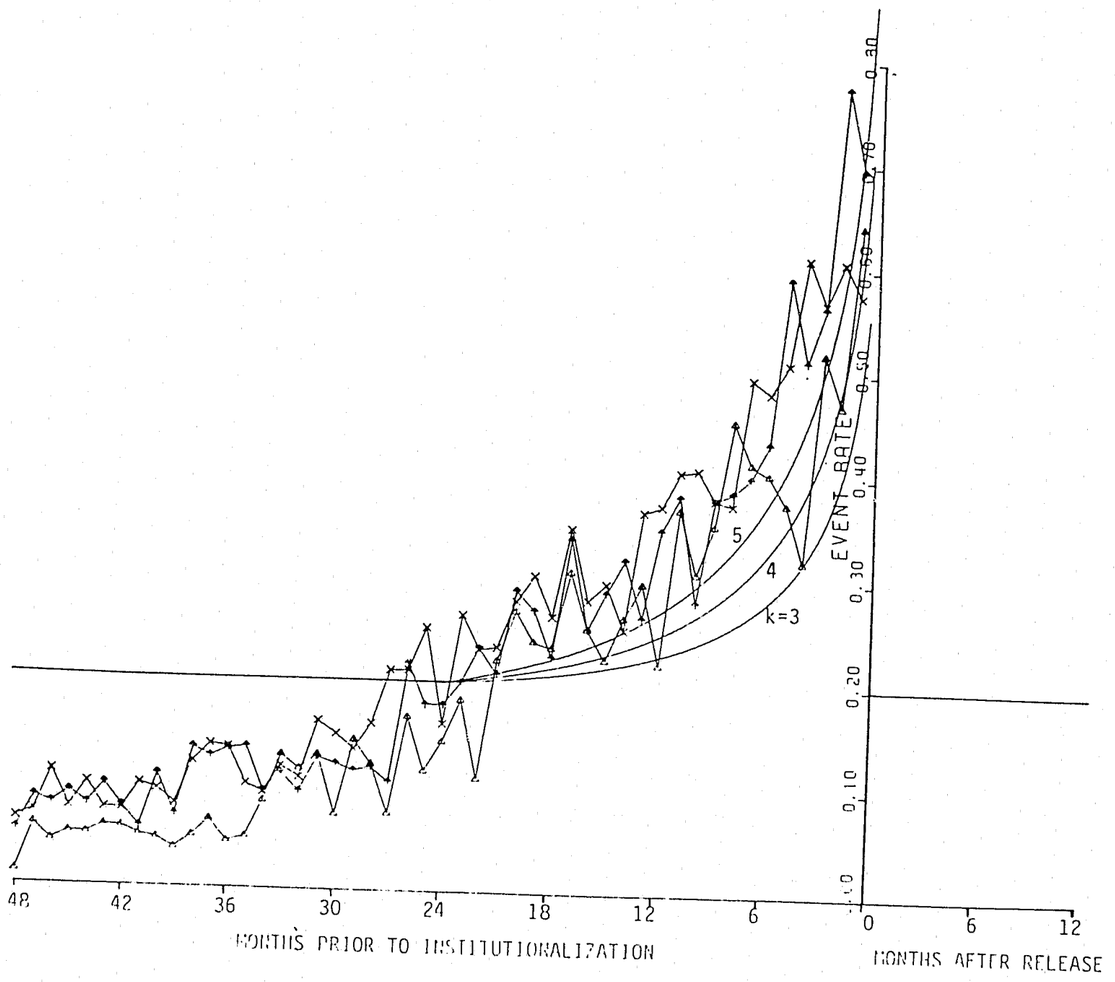


Figure 4-2 Effect of the Selection Artifact

next delinquent act that finally results in a conviction, i.e., a formal disposition. In other words, six delinquent acts (and these are only the ones we know of) have taken place in this scenario before formal intervention occurs. If the youth had terminated his delinquent career at an earlier point (i.e., with five or fewer acts) he would not have been subject to formal intervention.

A non-mathematical description of the selection artifact is given in Maltz et al. (1980). Basically, the artifact is generated because the selection of individuals for the cohort (i.e., by judges) is conditioned on the individuals' prior records. Those without a high rate of delinquent activity in the recent past are weeded out and do not appear in the cohort, so only those who have a period with a high rate of delinquent activity are included in the cohort. Thus, it is the judges who decide who will be members of the cohorts (experimental and control), and not the evaluators. Evaluators can compare the performance of the experimental and control groups after intervention, or they can attempt to determine the decision rule used by judges in selecting individuals for intervention, but they cannot compare delinquent activity before intervention with delinquent activity after intervention and expect to obtain valid findings.

This criticism of the "suppression effect" has in turn been criticized (Murray and Cox, 1676). The primary fault found with this analysis is that it relies too heavily on the assumption that offenders have a constant and random arrest rate. It is true that delinquents' arrest rates are not constant but are time-dependent: they rise until age 16 and from then on diminish (Wolfgang, Figlio & Sellin, 1972). But the analysis does not rely on this assumption: in fact, the point is that the steep rise of Figure 4-1 can be generated even with a constant arrest rate, not because of a constant arrest rate.

The assumption of randomness is more essential in explaining the artifactually high arrest rate just prior to institutionalization, which we will justify shortly. Murray and Cox (1979: 46) also make an assumption about the characteristics of offender activity, "that observed crime rate (arrests) is a valid proxy measure of real crime rate." In discussing the possibility of a regression artifact they refer not to the arrest rate, but to the crime rate.[2] But since they only have data on police contacts, they implicitly assume that delinquents' offense rates follow the same pattern as police contact rates.

If every offense resulted in a police contact there would be no difficulty with this assumption. But we know that this is not so; Boland and Wilson (1978) reviewed recent research which suggests that about one in five juvenile offenses results in a police contact. But even in this case it might be that the same pattern exists for offense rates and police contact rates -- for example, if the fifth, tenth, fifteenth, etc., offenses are the only ones that result in police

contacts. But regularity of this sort, which is required if police contacts are to be the proxy for offenses, is a very restrictive assumption and not likely to be the case (unless there is collusion or cooperation between offenders and police). It is more likely that the police contacts are not deterministically related to offenses: one would expect an occasional long run of offenses with no police contact, and an occasional period when police contacts bunch up. If a juvenile is selected for intervention due to a high rate of police contacts in the recent past, as we posit and as Terry (1967) has shown empirically to be the case, it may be due entirely to his having been unlucky recently rather than to a higher than average offense rate. It is this randomness that explains the artifactually steep rise in Figure 4-1 just prior to intervention. (An additional mathematical explanation is given in the last section of Chapter 7, below.)

The actual situation is probably somewhere in between the two extremes. Part of the "suppression effect" may be due to the treatment and the rest artifact. The part attributable to the artifact would be directly related to the ratio of arrests to offenses; the greater the proportion of offenses gotten away with by the offenders, the greater the contribution of the artifact.

Policy Implications

The two responses, "Nothing works" and "Anything works", are

attractive to legislators and other correctional policy-makers. Many people look to legislation for solutions to behavioral problems, and these answers point in specific legislative directions. Unlike so many other research results, they are definitive answers which can be easily understood and readily applied. They also appeal to those legislators and correctional administrators who take a hard-line approach, who feel that too much attention and care are given the offender and not enough are given the victim.

But the first answer oversimplifies a complex set of issues, and the second answer is not as definitive as it initially appeared to be. A seemingly equivocal answer ("Under these conditions one can expect this result thirty percent of the time") or the researcher's all-too-frequent response ("We need to do more research") may not be considered helpful, but with our present state of knowledge these are the only answers that can be given with any degree of confidence.[3] A well-planned program of research is needed to provide more definitive answers.

An important factor in planning and coordinating a correctional research program is ensuring that the results of different studies within the program are comparable. Similar settings and standardized operational definitions and analytic techniques are required. In succeeding chapters we discuss the way outcomes are affected by the criminal justice setting, by variations in the way recidivism has been

operationally defined, and by ways in which its definition might be standardized. Subsequent chapters review methods which have been used to analyze recidivism data, analytic techniques which have been recently developed, and examples of the use of these techniques in evaluating correctional programs and in understanding the recidivism process.

NOTES

1. A new program is often oversold by its advocates, who see little prospect of implementing it unless it is billed as a major breakthrough.

2. For example, they state: "We reject the regression artifact as an explanation....because it postulates a constant mean crime level with random fluctuations whereas the data indicate a rising mean." (Murray & Cox, 1979: 85). But we actually postulate a constant mean arrest level, and their data indicate nothing about crime levels.

3. Upon hearing a number of experts testify before his committee, all to the tune of, "On the one hand, ...; but on the other hand, ...", a senator was heard to say with exasperation, "I wish we had some one-handed scientists!" Unfortunately, the problems we face are not necessarily one-handed problems.

CHAPTER 5

CORRECTIONAL ORGANIZATION AND RECIDIVISM

In the previous chapter an example was given of how the organizational context of a study (i.e., the juvenile justice system) may have affected the measurement of its outcome. In this chapter we describe the effect of other organizational aspects of the criminal justice system on the measurement of correctional outcome. These organizations include the criminal courts, in their sentencing role and in their supervision of probation; correctional agencies, from halfway houses to work-release programs to prisons, and the correctional programs they offer; and parole boards and agencies, in determining the conditions of release and in their monitoring parolees' behavior to see that they hold to the conditions of release. Differences in these organizations limit direct state-to-state comparisons of correctional outcome.

We do not plan to study all of these organizations in depth and describe their impact on correctional outcome. Instead, in this chapter we describe the characteristics of one process -- parole -- and show how legal and organizational factors can influence measures of recidivism. Parolees constitute a significant proportion of prison

releasees. In addition, knowledge of their characteristics and of their post-release behavior is more readily available than for other releasees, since parole agencies are required to keep records on all of their clients. But state-to-state comparisons of parole outcomes (see Chapter 9) are meaningless unless one understands how state parole agencies differ and how interactions among all parts of the correctional system affect who get paroled.

The Parole Process

Sutherland and Cressey (1970: 584) provide this definition of parole:

> "Parole is the act of releasing or the status of being released from a penal or reformatory institution in which one has served a part of his maximum sentence, on condition of maintaining good behavior and remaining in the custody and under the guidance of the institution or some other agency approved by the state until a final discharge is granted."

There are many variants to this basic definition. One can view parole or conditional release as a process with a number of stages and actors responsible for the process in each of its stages. The stages in a typical parole process are

- sentencing as a means of defining parole eligibility,
- conditional release from a sentence, and
- supervision and discharge from or revocation of parole.

The actors in each of these stages are

- o courts
- o parole and pardon boards
- o parole field agencies and supervising agents.

The parole process is thus fully defined by a description of its stages, a description of the role of the actors in each of the stages, and a description of the relationships among actors in each of the stages. We may consider the following description as typical of the parole process in many states.[1]

Sentencing. The parole process is initiated in a sentencing court, the first stage. By setting minimum and maximum sentences, the court defines the time of parole eligibility. In general, a sentenced offender becomes eligible for parole

- o after serving the minimum time of the sentence (less time off for good behavior)[2] or twenty years, whichever is less -- in some states this this includes life sentences;
- o or after serving one-third of the maximum term (less time off for good behavior) or twenty years, whichever is less.

The sentencing court thus sets a time after which the sentenced offender is eligible for parole. In some cases the sentencing court may so restrict this time that, in effect, it takes away the parole decision-making function from the parole board.

Conditional release. The second stage of the parole process consists of a conditional release or parole decision by the state parole board. The board's decision-making task has four aspects:

- o deciding whether to release,
- o deciding the time of release,
- o deciding the conditions of release, and
- o deciding the time of discharge from supervision.

The decision in each case must meet specific legal criteria including an explanation or justification that is clear enough to permit its appeal to a court. Each decision must also be based on or must acknowledge data collected from other actors. The sentencing court and the state department of corrections furnish documents and reports related to each decision. For the most part, however, the interaction of these two actors with the board at this stage is pro forma.

Parole supervision. The third stage of the parole process consists of a period of parole supervision, terminated either by discharge from supervision or by revocation of parole. A parole officer is assigned to each case, and the board interacts with the parole officer in two ways. First, the officer is charged with enforcing the general conditions as well as any case-specific conditions of the release. For example, the parole board may release a person conditional upon his pariticipation in a special treatment program; the parole officer must enforce this special condition. Second, the

parole officer must participate in the decision to discharge the parolee from supervision or to revoke his parole. With respect to a discharge from supervision, the parole board continuously evaluates parolee behavior for a period of time to determine whether discharge is warranted. At the end of this period, given satisfactory performance and on the recommendation of the parole officer, the board issues a discharge order which operates as a commutation of sentence.

Of course, parole officers must report to the parole board all behaviors which appear to violate the conditions of release. The key word is "appear". In those cases in which a new crime is involved there will be little question as to whether the behavior violates the conditions of release. But in other cases the behavior may be interpreted ambiguously, so the validity of the charge must be tested. Differences in interpretation by different parole officers will result in different de facto revocation criteria.

Parole officers may have full police powers in the arrest and retaking of parolees. While the parole officer must request a warrant, a parolee can be detained pending issuance of the warrant.

Once the warrant has been issued, it must be heard by an officer designated by the parole board to determine whether there is cause for a revocation hearing. The hearing officer is usually an employee of the department of corrections or the parole board, so the charge, issuance of warrant and preliminary hearing on the warrant are all handled internally.

If the charge is upheld, a revocation hearing must be held by the parole board. On the basis of this hearing the board may decide either to revoke parole or to continue parole with or without modification of the conditions. The board may decide, for example, that the original conditions were too restrictive or unrealistic and may continue parole with looser restrictions. Or the board may decide that the original conditions were too loose or were not explicit enough and may continue parole with tighter conditions. If the board decides to revoke parole, however, the offender may be required to serve the remainder of his sentence (although a second conditional release is not precluded).

When a new crime is involved the parole officer in charge would be notified by the police, given an in-state arrest. Once this occurs, negotiations may take place between the police, prosecutors and parole officers. The charge may be dropped, may be reduced to a technical violation with subsequent repercussions or prosecution may ensue with a post-conviction sentencing option of revocation or return to parole. There are thus more alternatives available to the system in the handling of alleged repeat offenders who are parolees, because processing options present greater latitude due to the involvement and decision-making power of another organization -- the parole system. The only alternative available upon arrest and prosecution for a person mandatorily released is whether or not to proceed. Given that the al-

leged offender has served time in prison, it is likely that he would be formally processed. Thus, the same behavioral action does not always lead to the same disposition.

There is usually a clear distinction in the parolee's record between parole revocation for a technical violation of the conditions of release (e.g., failing to continue in a drug treatment program) and revocation in lieu of prosecution for a new offense. (The Uniform Parole Reports system makes such a distinction in its data collection procedures -- see Figure 9-12).

As can be seen from this description, each of the three stages -- determining parole eligibility, determining release conditions and supervising parolees -- is associated with a decision by one of the three actors -- the court, the parole board and the parole agency. The relative freedom enjoyed by each of the actors in each stage in turn is associated with the interaction of other actors at that stage. For example, the court may restrict the decision-making function of the parole board by limiting parole eligibility. And the parole board may restrict the decision-making function of the parole agency by setting conditions that are too restrictive or not restrictive enough.

Although this description of the parole process is quite general, there is still considerable variation in how parole is defined. The variation is due primarily to the limits on decision-making discretion placed on the actors by state legislatures. We may summarize the differences in the following way:

- o Sentencing is set by the court, but the degree of latitude in sentencing varies considerably from state to state. We may consider three general types of sentencing:
  - o Indeterminate sentencing: the sentencing court has wide discretion in the imposition of a sentence. Theoretically, it may impose prison terms of any length or range of lengths within broad guidelines; for example, courts have been known to impose sentences of 1000-3000 years on offenders, thus adhering to the state guideline that the minimum sentence be one-third the maximum sentence.
  - o Modified indeterminate sentencing: the sentencing court may impose a sentence that is charge-dependent, but the court may select a range from within this category. For example, if a certain charge is liable to a Class B penalty for which the maximum sentence may be set between five and ten years, the judge may select the maximum sentence anywhere between these two figures. If it is set at six years, and state law sets the minimum sentence at one-third the maximum, the offender leaves the courtroom knowing that he must serve at least two years and may serve at most six years -- less any "good time"[2] he may accrue while in prison.

o Determinate sentencing: the sentencing court imposes a sentence based primarily on the charge.[3] Latitude of the judge is limited to adding to the sentence if there are aggravating circumstances or reducing the sentence if there are mitigating circumstances. A limit is placed on the change in sentence length due to aggravation or mitigation; thus, if a certain charge is liable to a Class B penalty, which is a nominal five years but may be set anywhere between four and six years, the offender leaves the courtroom with a single number (between four and six years) which is his sentence (again, less good time). An additional feature of determinate sentencing is that the legislature may prescribe the circumstances which are allowed to be considered in aggravation or mitigation, and the judge may be required to document the circumstances considered in the determination of sentence length.

o Conditional release is related to the type of sentencing alternatives available to a state.

  o Indeterminate sentencing: Parole eligibility may be set by the court or the parole board. When this is the situation, one often finds prisoners attempting to curry favor with the authorities: "If joining Jaycees or an AA program or going to chapel is what gets me out sooner, that's what I'll do."

  o Determinate or modified determinate sentencing: The date of release is known by the offender when he is sentenced: the minimum sentence length, less any good time he may accrue. In some cases additional conditions may be imposed on the offender. Mandatory release may occur under determinate sentencing, in which case there are no post-release conditions to which the ex-offender must adhere. It may also occur under indeterminate sentencing if parole is denied and the prisoner serves his full sentence, i.e., "maxes out."

o Parole supervision does not vary to any great extent among the states. The primary difference is whether the supervision of parolees is directed by the parole board per se or by an independent agency working cooperatively with the parole board.

With regard to parole outcome, however, it appears that the structural status of supervision within the parole process makes little difference. What state-to-state differences that had existed in the past have been greatly reduced or eliminated by Supreme Court decisions in Morrissey v. Brewer and Gagnon v. Scarpelli. These decisions accorded the parolee due process rights in revocation procedures, such

as the right of the parolee to notification of the facts of the case, to be heard on his own behalf, to cross-examine witnesses, and to receive a written statement of the final decision and the reasons for it (Merritt, 1980). While the decisions in Morrissey and Gagnon leave some discretion to the states, the residual discretion is relatively small. For all practical purposes, the only formal difference in the parole process among the states is in the extent of authority vested in the courts and parole boards.

The Parole Process and Recidivism

The locus of authority for sentencing and release, as described in the last section, does not affect recidivism statistics directly. Yet its impact is substantial when the entire criminal process is considered.

Sentencing and Sentencing Alternatives. In recent years the trend has been for legislatures to limit the discretion of courts and parole boards. Thus, there has been a general shift toward determinate sentencing, vested good time, parole eligibility determined by time served rather than program participation; i.e., toward that legislative package known as "the justice model" (Fogel, 1979).

But discretion is not necessarily reduced by this legislative action; it may just shift to another level (Altschuler, 1978). For example, without determinate sentencing a prosecutor is able to drive

a harder (plea) bargain if the assigned judge is known to favor long sentences. With sentences fixed by law, it may simply mean that plea bargaining is based only on the charges rather than on the sentencing philosophy of the judge as well. So the prosecutor effectively determines the sentence given the offender.[4]

Another discretionary aspect of the sentencing process is in the type of correctional program open to an offender. The use of probation, work release, halfway houses, community treatment centers and other possibilities less severe than prison varies considerably among the states. But as Morris (1974: 11) has noted, the use of alternatives to prison results in "reducing the intensity and severity of control but increasing the numbers under control." A state which makes extensive use of these alternatives would be expected to have a higher recidivism rate for its parolees than would a state which does not use these alternatives greatly: the offenders considered to be "better risks" are given these alternatives, thus increasing the average risk level of those going to prison. Similar "skimming" of the lower-risk offenders would be manifested when comparing the recidivism rates of parolees to those denied parole.[5]

In comparing programs across states, therefore, care should be taken to ensure that the populations under study are similar. One means of doing so is to consider the proportion of people in each state sentenced to the various alternatives. If the proportions are

quite different comparisons are probably not very instructive; unfortunately, even if the proportions are similar one cannot be sure that the populations are also similar.

Release on parole. Parole boards have been criticized for arbitrariness in determining who is released (e.g., von Hirsch, 1976; Fogel, 1979). But another aspect of the release decision has not been given as much attention -- the timing of release. Eck (1979) has shown how, in one state, the number of people paroled in any given month is strongly correlated with the number of people sentenced to prison in that month. This policy serves to stabilize prison populations, which may partially explain why prison populations seemed to have a "natural" level in eighteenth-century France (Maltz, 1977: 32) and in twentieth-century Canada, Norway and United States (Blumstein & Cohen, 1973).

The impact of this policy would be felt in comparing cohorts of parolees from year to year. A surge in the number of persons sent to prison (perhaps reflecting new legislation, an increase in the number of judges, or a new prosecutorial policy toward plea bargaining) would cause an increase in the number of prisoners paroled. The only way for the parole board to accomplish this would be for them to lower their standards for release, resulting in a poorer-risk cohort for that year (at least, poorer in the board's estimation).

Parole supervision. Although parole officers do not set policy,

they do carry out the policies set by their supervisors. This may, for example, result in a strict revocation policy (say, in the aftermath of a publicized crime committed by a parolee) or a very lenient revocation policy (in response, say, to prison overcrowding). But "parole supervision" is not a constant entity; parole officers do not necessarily carry out their duties in the same way. In recounting the experiences of California's Special Intensive Parole Unit (SIPU) program, Conrad (1978) noted:

> "In Oakland, for example, the SIPU agent was an irrepressible enthusiast who kept his office open until late hours at night to dispense advice to, and to conduct bull sessions with any parolee who cared to happen in, as most of his caseload seemed to enjoy doing. His violation rate was extraordinarily low, and I never saw any reason to believe that there was a special ambience in Oakland which favored parole success. Across the bay in San Francisco the SIPU agent was an enthusiast of a different stripe. He liked to rise in the small hours of the morning so that he could descend on unemployed parolees and remind them that early birds get the available worms and slug-a-beds do not. How he managed to conduct these sunrise raids on his charges without dismemberment of his person I have never understood, but his parole violation rate was high, even after he was convinced of the unwisdom of the strenuous counseling technique he had adopted."

Variation in outcomes, then, is not only a function of legislative or policy differences; the personal attributes of the actors also influnece recidivism rates.

Another factor influencing program outcomes lies somewhere in between agency policy and personal proclivity. This might be termed the agency's "style". Wilson (1968) identified three dominant styles of policing -- legalistic, watchman and service styles. Similar styles doubtless are characteristic of correctional agencies. However, studies of correctional agencies tend to be of single entities (e.g., Jacobs, 1978; McCleary, 1978) rather than comparative or cross-sectional. And police departments are geographically compact while correctional organizations are dispersed throughout a state. Therefore, a single style may not be so dominant in a correctional organization as in a police department; there may be major differences between rural and urban (or upstate-downstate) parole supervision styles, for example.

Empirical Variations

Not all factors that have an impact on recidivism can be found by reviewing agency policy statements or by reading the literature. Some can only be found by making site visits. As part of our research effort eleven state correctional and parole agencies were visited[6] to determine how their characteristics and policies might affect the

outcome measures used to study correctional programs. Our brief visits did not enable us to analyze an agency's "style", but we were able to obtain insights into other factors that affect recidivism statistics. We obtained information on the outcome measures used, the populations followed up, the events which define recidivism, and the length of follow-up times, as well as organizational factors which work to produce variations in recidivism rates.

Outcome measures used. All routine evaluations used recidivism as the criterion variable. No measures of success were calculated. Reasons given for this were "it is too difficult to collect such data", "it violates privacy rights", "it is not part of our mission", and "the cost of data collection is too high for measures which have little bearing on policy". Thus, the evaluations are failure-based. Even given more research money, many officials contended it would not be used to develop success measures. Some officials reported that recidivism data are collected only to satisfy other agencies which demand such data, such as the Law Enforcement Assistance Administration (LEAA), the state legislature, or the National Council on Crime and Delinquency, or to forecast prison populations.

Populations followed up. Some variation was found in the types of populations followed up. Four states followed up only parolees, while three followed up all persons released, including parolees, mandatory conditional releasees, and mandatory releasees. Another state studied

all but those mandatorily released while still another followed up a cohort of parolees and a cohort of those mandatorily released.

Naturally, such differences will have an effect upon observed failure rates. Since parolees are considered better risks upon release, one would anticipate that ceteris paribus states which follow up only parolees would have a lower recidivism rate than states which follow up those released under any status.

Parolees, however, are subject to more behavioral constraints than those mandatorily released. A violation of the technical conditions of parole can result in the revocation of parole and consequent readmission to prison. Parolees may also be under closer and more systematic observation. The greater observation of parolees, then, may work in the opposite direction; that is, one might also expect parolees to have a higher recidivism rate.[7]

Recidivating events. Events defined as incidents of failure showed great convergence. Six states examined considered return to prison as the only indicator of failure. Thus, absconders, technical violators, and those convicted of new crimes would only be documented as failures if these actions led to subsequent recommitment. Furthermore, only one state counted returns to prison if they occurred in another state.

Some states' definitions approach more rigorously the pragmatic conception of a failure; they extend the definition of failure beyond

return to prison. One state included a new major conviction with subsequent continuation on parole as a failing event, while another included a jail sentence of thirty days or more as an indicator of failure.

Finally, one state utilized a fairly comprehensive definition of recidivism, including: parolee-at-large for six months or longer; death in the commission of a crime or from drug overdose; any new conviction which resulted in one of six types of sanctions; felony arrest and charge, guilt admitted, no further trial; and, return to prison for felony conviction in any state.

Absconsion. The interpretation of absconscion in recidivism measurement varies across states. Officials' estimates of the proportion of parolees who abscond ranged from two to ten percent. One state considers all officially recorded absconders as failures, while another considers them failures if they have been in such status for at least six months or have an outstanding felony warrant for their arrest. The remaining states consider them non-recidivists, or "successes", if they are not returned to prison in-state during the duration of the follow-up.

Event time. The date of occurrence of a recidivating event is the time at which the recidivist and the system which holds the information interact. Thus, in the case of return to prison, the date the individual is "on the books" as having returned to prison is the date

of the recidivating event. Events are not traced back to the time of the actual act (e.g., arrest) which caused subsequent system interaction. Recidivism is calculated using information relating to reimprisonment, and only that information which is of ultimate concern to the correctional agency may be formally documented. In special evaluations, however, FBI or state "rap sheet" data may be used to define failure, and date of arrest would be used as the event time.

Follow-up period. The range of maximum followup periods for the states visited was from one to four years, the most common being one year. Of course, not all of the parolees are followed up for the four years in those states which track parolees for that length of time. Only those who are on parole for that length of time are followed up for four years. Thus, recidivism statistics for the fourth year would be biased because they would reflect the behavior of only that subset of parolees for whom a four-year parole is considered necessary.

Organizational factors in reporting failure events. Parole officers use considerable discretion when deciding whether to report technical violations and institute revocation procedures (McCleary, 1978; Lerman, 1968). One parole official interviewed admitted: "There is great pressure on parole officers not to return parolees due to the overcrowded situation of our prisons." Thus, the capacity of a state to incarcerate, the number of "free beds", could have an eventual effect upon failure rates, especially those based upon return to prison.

Other factors also affect the underreporting of parole violations. According to officials, the amount of paperwork required to process violations could have an effect upon the stringency of reporting, especially when considered in relation to officer-clientele ratios. Departmental policy on revocations, as mandated by internal and external organizational constraints and contingencies, will also have an effect upon eventual rate of return. The Supreme Court's decisions related to parole have made it more difficult for parole officers to revoke parole. Whereas before considerable discretion was given to parole officers to revoke, some officials say that now such discretion operates in the opposite manner -- considerable discretion not to revoke.

Agency perspective. From the standpoint of correctional agencies, the post-release behavior of individuals who were once in their custody is not their responsibility. This is especially true in times of tight budgetary constraints, when maintaining livable conditions in institutions or keeping parole caseloads at a reasonable level takes priority over studying recidivism.

Agency officials are also closer to the source of data used to evaluate programs, and are therefore much more cautious in making inferences about programs based on such data. One official explained that evaluations which relied on data from parole officer reports were discontinued because "we came to consider them too biased to use,

because information collected on parolees was generally in narrative form and primarily reflective of the parole officer's personal attitude toward his client."

Thus, variation in parole organization policy and practice, as well as variation in the types of releasees who are followed up, will result in varying observed rates of recidivism.

In this chapter we have described characteristics of correctional organizations that affect recidivism. Differences in the characteristics may create differences in the population under study, in the way the program is conducted, or in the type of data collected. But another significant factor limiting comparisons is the lack of consistency in defining recidivism. This problem is taken up in the next chapter.

NOTES

1. The description of parole is based on the Illinois parole process as it existed prior to 1978, when determinate sentencing legislation was enacted. Many states still have similar organizational structures for parole.

2. "Good time" is the time subtracted from a prisoner's sentence for good behavior while in prison. One characteristic of the justice model of corrections (Fogel, 1979) is the vesting of good time so that all of it cannot be taken away, once accumulated, for subsequent infractions of institution rules.

3. The offender's prior record may also affect sentence length. For example, a life sentence may be mandatory if an offender has a record of three prior felony convictions. But these conditions are also non-discretionary, i.e., determinate.

4. Plea bargaining is not easy to eliminate, especially in jurisdictions with crowded dockets. But even in jurisdictions with less crowded dockets it seems to survive all attempts to eradicate it. In Alaska, for example, plea bargaining was outlawed by the state

legislature. But according to a recent report the practice still continues (Rubinstein & White, 1979).

5. For example, Martinson and Wilks (1977) noted the higher recidivism rate of prisoners denied parole than of parolees. However, they misinterpreted this difference as demonstrating the effectiveness of parole rather than as being a consequence of the parole boards' selection of the best risks for parole.

6. The states visited were California, Florida, Georgia, Maine, Massachusetts, Michigan, Minnesota, Texas, Washington and Wisconsin.

7. Care should be taken to distinguish the two effects described here. The first effect is that the recidivism rate is expected to be lower because a population of parolees is expected to have a lower average offense rate than a population which includes those denied parole. The other effect is that closer observation of parolees may increase the probability of arrest for any given offense.

CHAPTER 6

RECIDIVISM DEFINITIONS

When "recidivism" is discussed in a correctional context, its meaning seems fairly clear. A person is considered to have recidivated if his (criminal) behavior has not been improved by the correctional intervention. The word is derived from the Latin "recidere," to fall back. Thus a recidivist is one who, after his release from custody for having committed a crime, is not rehabilitated. Instead, he falls back, or relapses, into his former behavioral patterns: he commits more crimes.

This conceptual definition seems quite straightforward, but like most conceptual definitions its operationalization into something one can measure is not so simple. The information on which measurement of recidivism is based is rarely complete; and even when it is complete there is no consistency in the way the data are analyzed: there is "considerable variation in the way recidivism is measured" (National Advisory Commission, 1973: 512).

In this chapter we discuss variations in the way recidivism is defined. We will begin by assuming that all conceivable information one would need is available, then show how more realistic assumptions about information affect the definitions. Our recommendations concerning the measurement of recidivism conclude this chapter.

Complete Information

By "complete information" we mean that every crime committed by the individuals under study is known to the evaluators. But even under this unlikely condition there is a problem, since the word "crime" covers a lot of ground.

For example, a child molester may be arrested, convicted, and sentenced to a correctional program specifically designed to treat such offenders. Upon release he may have actually been corrected so that he no longer molests children. If he then turns to armed robbery or to forgery, should we consider him a recidivist? In one sense he is a recidivist since he committed another crime. However, the crime is of an entirely different nature; the original harmful behavior has ceased and a different set of harmful behaviors has manifested itself. Should we consider a person a recidivist if he crosses over from one crime type to another?

Of course, it is possible to label a person a recidivist only if he commits the same crime type for which he was originally convicted. However, this implies that all offenders are specialists, which is contradicted by the available evidence (Petersilia et al, 1977). Furthermore, the more categories of recidivism we generate, the less information we will have about each category. One might consider using three categories of crime: property crime (including robbery[1]), personal crime, public order crime, and white-collar crime. But there

are no clear dividing lines among these categories, since many crimes, e.g., arson, can cross the boundaries.

To a great extent the nature of the correctional program dictates the definition of recidivism. In the example given above the individual would not be considered a recidivist if one were evaluating a program aimed at modifying the behavior of child molesters. Thus, both crime type and program type must be considered in defining recidivism.

Complete Criminal Justice System Information

Now let us assume that we have the best of all possible realistic worlds. We do not have information about all crimes committed by individuals in the cohort, but we do have complete information about all of their transactions with the criminal justice system: arrests, indictments, prosecutions, convictions, and sentences. This goal is achievable -- the Law Enforcement Assistance Administration is providing funds for states to develop Offender-Based Transaction Systems (OBTS), which are designed to collect such data. A number of states currently have substantial OBTS data collection and analysis capability.

The fundamental question we must now address is, To what extent can an individual's criminal record be used as an indicator of his behavior? We know that it has major weaknesses; since it is the only

indicator we have (other than, possibly, self-reports) our goal is to understand these weaknesses and use the criminal record in a way that minimizes their effect on the behavioral measure used.

One problem we may face is that of determining who is a recidivist and who is a first offender. Many states have laws prohibiting disclosure of an individual's juvenile record so that youthful indiscretions and delinquent activities do not haunt him for the rest of his life. This policy transforms everyone into a first offender upon reaching the age at which he can be tried as an adult, regardless of his past record. However, juvenile records can often be used for research purposes if appropriate safeguards are taken to insure confidentiality.

It is important to determine <u>when</u> the behavior occurred. Normally, the date of arrest (or of violation, for parolees and probationers) is the only indicator of time. If arrest date coincides with offense date this is a good indicator; but this is the exception rather than the rule. For the purposes of the analysis we describe in later chapters, however, it is sufficient if offense and arrest occur within one month of each other. It is not unreasonable to assume that is ordinarily the case.

A more important problem in defining recidivism is whether to use a raw <u>arrest</u> (one not necessarily followed by a <u>conviction</u>) as an indicator. Were we dealing with named individuals there could only be

one appropriate answer: an arrest must be followed by a conviction before it can be used as an indicator of behavior; a person should not be <u>assumed guilty</u> merely because he has been arrested. But we are not dealing with named individuals; rather, we are dealing with statistical descriptors of <u>cohorts</u>.

From the social scientist's standpoint the primary consideration is to use the available data to develop the most appropriate indicator, the one that is closest to what we think of as recidivism. In practical terms, this boils down to a choice between using raw arrest data or using arrests only if followed by conviction.[2]

The arguments against using raw arrests are based on the fact that standards for arrest are much less rigorous than for conviction. Probable cause is sufficient to arrest an individual; proof beyond reasonable doubt is needed to convict him. Furthermore, the arrest of a person released from prison (i.e., known to have been an offender) is much more likely than the arrest of that same person had he no prior record. For example, let us suppose that a person convicted of child molesting has actually been rehabilitated. This does not make him immune from arrest; he may be subject (and subjected) to arrest frequently, whenever a child is molested anywhere nearby. An arrest of this type should not be an indicator of recidivism.

Arrests are used for other purposes than detaining those known to have committed crimes. It may be that the police have a policy of

CONTINUED

1 OF 4

*Hansen, Hans Alois and Karl Teilmann. "A Treatment of Criminal Alcoholics in Denmark," (1954). Cited in Lipton et al (1975).

Harris, Carl M., Ali R. Kaylan, and Michael D. Maltz. "Advances in Statistics of Recidivism Measurement." To appear in J.A. Fox, Ed., Mathematical Frontiers in Criminology. New York: Academic Press, 1981.

Harris, Carl M. and Jay Mandelbaum. "Asymptotic Behavior of Maximum-Likelihood Estimators of Mixed Weibull Parameters." Project Memo CMH1-UICC-DOJ for the Center for Research in Criminal Justice, University of Illinois at Chicago Circle, 1979.

Harris, Carl M. and Soumyo Moitra. "Improved Statistical Techniques for the Measurement of Recidivism." Journal of Research in Crime and Delinquency 15 (1978): 194-213.

Harris, C. and S. Moitra. "On the Transfer of Some OR/MS Technology." Interfaces 9 (1978): 78-86.

*Havel, Joan. "A Synopsis of Research Report No. 10 SIPU Phase 4 - The High Base Expectancy Study," (1963). Cited in Lipton et al (1975).

*Havel, Joan. "Special Intensive Parole Unit - Phase Four: 'The Parole Outcome Study,'" (1965). Cited in Lipton et al (1975).

Haybittle, J.L. "A Two-Parameter Model for the Survival Curve of Treated Cancer Patients." Journal of the American Statistical Association (1965): 16-26.

Herman, Robert and Ilya Prigogine. "A Two-Fluid Approach to Town Traffic." Science 204 (1979): 148-151.

Hoffman, P.B. and J.L. Beck. "Parole Decision-Making: A Salient Factor Score." Journal of Criminal Justice 2 (1974): 195-206.

Hoffman, Peter B. and Barbara Stone-Meierhoefer. "Post Release Arrest Experiences of Federal Prisoners - a Six Year Followup." United States Parole Commission Research Unit, Report 17, December, 1977. Also

published in Journal of Criminal Justice 7 (1979): 193-216.

Hoffman, Peter B. and Barbara Stone-Meierhoefer. "Reporting Recidivism Rates: The Criterion/Followup Issue." United States Parole Commission Research Unit, March, 1980.

Hoge, Warren. "Brazil's Crime Rate Produces Lynchings." New York Times, December 2, 1979, Section 1, page 11.

*Hood, Roger. Homeless Borstal Boys: A Study of Their After-Care and After-Conduct, (1966). Cited in Lipton et al (1975).

Hopkins, Andrew. "Imprisonment and Recidivism: A Quasi-Experimental Study." Journal of Research in Crime and Delinquency 13 (1976): 13-32.

Inciardi, James. "The Use of Parole Prediction with Institutionalized Narcotics Addicts." Journal of Research in Crime and Delinquency 8 (1971):

Jacobs, James B. Stateville: The Penitentiary in Mass Society. Chicago: University of Chicagoo Press, 1978.

*Jacobson, Frank and Eugene McGee. "Englewood Project: Re-Education: A Radical Correction of Incarcerated Delinquents," (1965). Cited in Lipton et al (1975).

*Johnson, Bertram. "Parole Performance of the First Year's Releases, Parole Research Project: Evaluation of Reduced Caseloads," (1962). Cited in Lipton et al (1975).

*Johnson, Bertram. "An Analysis of Predictions of Parole Performance and of Judgments of Supervision in the Parole Research Project." (1962). Cited in Lipton et al (1975).

Joplin, G.H. "Self-Concept and the Highfields Program: Recidivists vs. Non-Recidivists." Criminology 9 (1973): 491-495.

Kaplan, E.L., and Meier, Paul. "Nonparametric Estmation from Incomplete Observations." Journal of the American Statistical Association (1958): 457-481.

Kassebaum, Gene, David Ward, and Daniel Wilner. Prison Treatment and Parole Survival: an Empirical Assessment. New York: Wiley, 1971.

Kantrowitz, Nathan. "How to Shorten the Follow-up Period in Parole Studies." Journal of Research in Crime and Delinquency 14 (1977): 222-236. (1977): 222-236.

Kaplan, Edward H. "Models for the Evaluation of Treatment Release Corrections Programs." Technical Report No. 162, Operations Research Center, Massachusetts Institute of Technology, May, 1979.

*Kawaguchi, Ray M., and Siff, Leon M. "An Analysis of Intensive Probation Services - Phase II," (1967). Cited in Lipton et al (1975).

Kelling, G.L., T. Pate, D. Dickman, and C.E. Brown. The Kansas City Preventive Patrol Experiment and Summary Report. Police Foundation, 1015 18th Street, N.W., Washington, DC, 1974.

Kelly, F.J. and D.J. Baer. "Physical Challenge as a Treatment for Delinquency." Crime and Delinquency 17 (1971): 437-445.

Kiernan, L. "Study Says Youth Crime Falls After Incarceration." Washington Star, Page A4, November 19, 1979.

Kitchener, Howard, Annesley K. Schmidt, and Daniel Glaser. "How Persistent Is Post-Prison Success?" Federal Probation 41 (1977): 9-15.

Kitsuse, J. and A. Cicourel. "A Note on the Use of Official Data." Social Problems 11 (1963): 131-139.

Klein, M.W. "Labeling, Deterrence, and Recidivism: A Study of Police Dispositions of Juvenile Offenders." Social Problems 22 (1974): 292-303.

Klockars, Carl B. "The True Limits of the Effectiveness of Correctional Treatment." Prison Journal 55 (1975): 53-64.

Klockars, C. B. "Letter to the Editor." Prison Journal 55 (1975): 65.

Kovacs, Frank W. "Evaluation and Final report of the New Start Demonstrtion Project," (1967). Cited in Lipton et al (1975).

Kozol, H.L., R.J. Boucher, and R.F. Garofalo. "The Diagnosis and Treatment of Dangerousness." Crime and Delinquency 18 (1972): 371-392.

*Kusuda, Paul H., and Babst, Dean V. "Wisconsin Base Expectancies for Adult Male Parolees: Preliminary Findings on the Relationship of Training, Work and Institutional Adjustment at Wisconsin State Prisn to Later Parole Violation Experience, 1958 and 1959 Releases." (1964). Cited in Lipton et al (1975).

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

LeClair, Daniel P. An Analysis of Recidivism Rates Among Residents Released from Massachusetts Correctional Institutions During the Year 1974. Massachusetts Department of Corrections. 1977.

Lerman, Paul. "Evaluative Studies of Institutions for Delinquents." Social Work, 13, 3, (July, 1978).

Lipton, Douglas, Robert Martinson, and Judith Wilks. The Effectiveness of Correctional Treatment: A Survey of Treatment Evaluation Studies. New York: Praeger Publishers, 1975.

Lloyd, Michael R., and George Joe. "Recidivism Comparisons Across Groups: Methods of Estimation and Tests of Significance for Recidivism Rate and Asymptotes." Evaluation Quarterly (to appear).

*Lohman, Joseph D., et al. "The Intensive Supervision Caseloads: A Preliminary Evaluation." (1967). Cited in Lipton et al (1975).

Maltz, Michael D. Evaluation of Crime Control Programs. Washington DC: U.S. Government Printing Office, 1972.

Maltz, Michael D., Andrew C. Gordon, David McDowall, and Richard McCleary. "Why Before-After Comparisons of Criminal Activity Should Not Be Used to Evaluate Delinquency Programs." Evaluation Review, April 1980.

Maltz, Michael D., and Richard McCleary. "The Mathematics of Behavioral Change: Recidivism and Construct Validity." Evaluation Quarterly 1 (1977): 421-438.

Maltz, Michael D., and Richard McCleary. "Comments on 'Stability of Parameter Estimates in the Split Population Exponential Distributions.'" Evaluation Quarterly 2 (1978): 650-654.

Maltz, Michael, D., Richard McCleary, and Stephen M. Pollock. "Recidivism and Likelihood Functions: Reply to Stollmack." Evaluation Quarterly 3 (1979): 124-131.

Maltz, Michael D., and Stephen M. Pollock. "Artificial Inflation of a Poisson Rate by a 'Selection Effect.'" Operations Research 28, 3, May-June (1980): 547-559.

Maltz, Michael D. and Stephen M. Pollock. "Bayes Estimates for Singly-Censored Data From a Mixed Failure Distribution" (forthcoming).

Marcus, Richard, and Saul Blumenthal. "A Sequential Screening Procedure." Technometrics 16 (1974): 229-234.

Martinson, Robert. "What Works? Questions and Answers about Prison Reform." The Public Interest (Spring, 1974): 22-54.

Martinson, Robert. "Letter to the Editor." Prison Journal 55 (1975): 65.

Martinson, Robert, Ted Palmer, and Stuart Adams. Rehabilitation, Recidivism, and Research. Hackensack, NJ: National Council on Crime and Delinquency, 1976.

Martinson, Robert and Judith Wilks. "Save Parole Supervision." Federal Probation 41 (1977): 23-27.

Marx, Gary, and Dane Archer. "Citizen Involvement in the Law Enforcement Process: The Case of Community Police Patrols." American Behavioral Scientist 15 (1971): 52-72.

Marx, Gary, and Dane Archer. "The Urban Vigilante." Psychology Today, January 1973, 45-50.

*Massimo, Joseph L., and Shore, Milton F. "The Effectiveness of a Comprehensive Vocationaly Oriented Psychotherapeutic Program for Adolescent Delinquent Boys." (1963). Cited in Lipton et al (1975).

*McCorkle, Lloyd W., Albert Elias, and F. Lovell Bixby. The Highfields Story: An Experimental Treatment Project for Youthful Offenders. (1958). Cited in Lipton et al (1975).

*McEachern, Alexander W., and Taylor, Edward M. "The Effects of Probation." (1967). Cited in Lipton et al (1975).

Meade, A. "Seriousness of Delinquency: The Adjudicative Decision and Recidivism - A Longitudinal Configuration Analysis." Journal of Criminal Law, Criminology and Police Science 64 (1973): 478-485.

Merritt, Frank S. "Parole Revocation: A Primer." University of Toledo Law Review 11, Summer 1980: 893-938.

McCleary, Richard. Dangerous Men. V. 71. Beverly Hills: Sage Publications, Inc., 1978.

McCleary, R., A.C. Gordon, D. McDowall, and M.D. Maltz. "How a Regression Artifact Can Make Any Delinquency Program Look Effective." In Sechrest et al., Eds. Evaluation Studies Review Annual. Vol. 4. Beverly Hills, Cal.: Sage Publications, 1979.

Miley, Alan D. "Stability of Parameter Estimates in the Split Population Exponential Distribution." Evaluation Quarterly 2 (1978): 646-649.

Miller, S.J. and Simon Dinitz. "Measuring Institutional Impact." Criminology 11 (1973): 417-426.

Minnesota Governor's Commission on Crime Prevention and Control. Evaluation Unit. Residential Community Corrections Programs in Minneosota. St. Paul: Minnesota Governor's Commission, November, 1976.

Moberg, David O., and Ericson, Richard C. "A New Recidivism Outcome Index." Federal Probation 36 (1972): 50-57.

Morris, Norval. The Future of Imprisonment. Chicago: University of Chicago Press, 1974.

Moseley, William and Margaret Gerould. "Sex and Parole: A Comparison of Male and Female Parolees." Journal of Criminal Justice 3 (1975): 47-58.

Murray, C.A. Statement Before the United States Senate Subcommittee on Juvenile Delinquency. April 12, 1978a.

Murray, C.A. Letter to the Editor. Criminal Justice Newsletter, 1978b, 9(16), 4-5.

Murray,. C.A. Interview by Garrick Utley, NBC TODAY program, November 16, 1979.

Murray, C.A. and L.A. Cox, Jr. "Juvenile Corrections and the the Chronic Delinquent." American Institutes for Research. Washington, DC, 1979.

Murray, C.A. and L.A. Cox, Jr. Beyond Probation: Juvenile Corrections and the Chronic Delinquent. Sage: Los Angeles, 1979.

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

National Advisory Commission on Criminal Justice Standards and Goals. (1973): Corrections. Washington DC, U.S. Government Printing Office.

*Narloch, R.P.; Adams, Stuart; and Jenkins, Kendall J. Characteristics and Parole Performance of California Youth Authority Early Releases." (1959). Cited in Lipton et al (1975).

Neithercutt, M.G., William H. Moseley, and Ernst A. Wenk. "Uniform Parle Reports: A National Correctional Data System." Report by the National Council on Crime and Delinquency Research Center, 1975.

New York Times. Editorial, June 19, 1978. Letter to the Editor, June 30, 1978.

*New York State. Division of Parole, Department of Correction. "Parole Adjustment and Prior educational Achievement of Male Adolescent Offenders, June 1957-June 1961." (1964). Cited in Lipton et al (1975).

Palmer, Ted. "Martinson Revisited." Journal of Research in Crime and Delinquency 12 (1975): 133-152.

Palmer. T. Correctional Intervention and Research: Current Issues and Future Prospects. Lexington, MA: Lexington Books, 1978.

Partanen, Juha. "On Waiting Time Distributions." Acta Sociologica 12 (1969): 132-143.

*Persons, Roy W. "Relationship between Psychotherapy with Institutionalized Boys and Subsequent Community Adjustment," (1967). Cited in Lipton, et al (1975).

Petersilia, J., P. Greenwood, and M. Lavin. "Criminal Careers of Habitual Felons. Report No. R-2144-DOJ. Santa Monica, Rand Corporation, 1977.

Pew, M.L., D.C. Speer, and J. Williams. "Group Counseling for Offenders." Social Work 18 (1973): 74-79.

*Pilnick, Saul, et al. "Collegefields: From Delinquency to Freedom," (1967). Cited in Lipton, et al (1975).

Philpotts, G.J.O., and L.B. Lancucki. Previous Convictions, Sentence and Reconviction. Home Office Research Study No. 53. London: Her Majesty's Stationery Office, 1979.

Pierce, G.L., and W.J. Bowers. "The Impact of the Bartley-Fox Gun Contrl Law on Crime in Massachusetts," Center for Applied Social Research, Northeastern University, Boston, April, 1979.

Press, S.J. Some Effects of an Increase in Police Manpower in the 20th Precient R-704-NYC. Santa Monica, The Rand Corporation, 1971.

Rawls, John. A Theory of Justice. Cambridge, MA: Harvard University Press, 1972.

Reed, David E. "Whistlestop: A Community Alternative in Crime Control." Unpublished Ph.D. dissertation, Department of Sociology, Northwestern University, Evanston, Illinois, 1978.

Rezmovic, Eva L. "Methodological Considerations in Evaluating Correctional Effectiveness: Issues and Chronic Problems." In Sechrest et al (1979), 163-209.

Rohatgi, V. An Introduction to Probability Theory and Mathematical Statistics. New York: Wiley, 1975.

Rubinstein, M.L., and T.J. White. "Plea Bargaining: Can Alaska Live without It?" Judicature 62 (1979): 266-279.

*Rudoff, Alvin. "The Effect of Treatment on Incarcerated Young Adult Delinquents as Measured by Disciplinary History." (1960). Cited in Lipton, et al, (1975).

Schmidt, Peter and Ann D. Witte. "Determinants of Criminal Recidivism: Further Investigations." Report to the North Carolina Department of Corrections on Grant 77-04-B07-1013. 1978.

Schmidt, P. and A. D. Witte. "Models of Criminal Recidivism and an Illustration of Their Use in Evaluating Correctional Programs." University of North Carolina, no date.

*Schwitzgebel, Robert, and Kolb, D.A. "Inducing Behavior Change in Adolescent Delinquents," (1964). Cited in Lipton, et al, (1975).

Sechrest, Lee, Susan O. White, and Elizabeth D. Brown, Editors. The Rehabilitation of Criminal Offenders: Problems and Prospects. Washington, DC: National Academy of Sciences, 1979.

Sellin, Thorsten and Wolfgang, Marvin E. The Measurement of Delinquency. Wiley: New York, 1964.

*Shelley, Ernest L.V., and Johnson, Walter F., Jr. "Evaluating an Organized Counseling Service for Youthful Offenders," (1961). Cited in Lipton, et al (1975).

Shinnar, R. and S. Shinnar. "The Effects of the Criminal Justice System on the Control of Crime: A Quantitative Approach." Law and Society Review 9 (1975): 581-611.

*Shoham, Shlomo, and Sandberg, Moshe. "Suspended Sentences in Israel: An Evaluation of the Preventive Efficacy of Prospective Imprisonment," (1964). Cited in Lipton, et al (1975).

Simon, Frances H. "Statistical Methods of Making Prediction Instruments." Journal of Research in Crime and Delinquency 9 (1972): 46-53.

Sokal, Robert R. "Classification: Purposes, Principles, Progress, Prospects." Science 185 (1974): 1115-1123.

Stamp, Sir Josiah. Some Economic Factors in Modern Life. London. P.S. King and Son, Ltd. 1929.

*Stanton, John M. "An Empirical Study of the Results of the Special Narcotics Project," (1956). Cited in Lipton, et al, (1975).

Stollmack, Stephen. "Comments on 'The Mathematics of Behavioral Change.'" Evaluation Quarterly 3 (1979): 120-124.

Stollmack, Stephen, and Harris, Carl M. "Failure-Rate Analysis Applied to Recidivism Data." Operations Research 22 (1974): 1192-1205.

*Sullivan, Clyde E., and Mandell, Wallace. "Restoration of Youth through Training: A Final Report." (1967) Cited in Lipton, et al (1975).

Sutherland, Edwin, and D. Cressey. Principles of Criminology. Philadelphia: Lippincott, 1970.

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

Turnbull, Bruce W. "A Note on the Nonparametric Analysis of the Stollmack-Harris Recidivism Data." Operations Research 25 (1977): 706-708.

U.S. Department of Justice, Law Enforcement Assistance Administration, National Criminal Justice Information and Statistics Service. Criminal Victimization in the United States 1974. Washington DC: U.S. Government Printing Office, 1977.

Van Alstyne, David J. and Michael R. Gottfredson. "A Multi-Dimensional Contingency Table Analysis of Parole Outcome: New Methods and Old Problems in Criminological Prediction." Journal of Research in Crime and Delinquency 15 (July, 1978): 171-193.

*Van Couvering, Nancy, et al. "One-to-One Project: Final Report." (1966). Cited in Lipton, et al (1975)

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

*Waldo, Gordon P. and Theodore G. Chiricos. "Work Release and Recidivism: An Empirical Evaluation of a Social Policy". Evaluation Quarterly 1 (February, 1977): 87-108.

Werner, Eric and Ted Palmer. "Psychological Characteristics of Successful and Unsuccessful Parolees: Implications of Heteroscedastic and Non-Linear Relationships." Journal of Research in Crime and Delinquency 13 (July, 1976): 165-178.

*Wilkins, Leslie T. "A Small Comparative Study of the Results of Probation." (1958). Cited in Lipton, et al (1975).

Wilkins, Leslie T. Evaluation of Penal Measures. New York: Random House, 1969.

Witte, Ann D. and Peter Schmidt. "An Analysis of Recidivism, Using the Truncated Lognormal Distribution." Applied Statistics 26 (1977): 302-311.

Wolfgang, Marvin E., Robert M. Figlio, and Thorsten Sellin. Delinquency in a Birth Cohort. Chicago: University of Illinois, 1972.

*Zivan, Morton. "Youth in Trouble: A Vocational Approach," (1966). Cited in Lipton, et al (1975).

Zimring, F.E. and G.J. Hawkins. Deterrence: The Legal Threat in Crime Control. Chicago: University of Chicago Press. 1976.

*Zumbrun, Alvin J. T., and Berry, John. "A Follow-Up Study of Men Released in 1954 from the State Reformatory for Males, by Reason of Parole and Expiration of Sentence." (1958). Cited in Lipton, et al (1975).