*A QUASI-EXPERIMENTAL RETROSPECTIVE EVALUATION
OF A PRISON POLICY CHANGE¹*

JOHN F. SCHNELLE AND J. FRANK LEE

MIDDLE TENNESSEE UNIVERSITY, RUTHERFORD COUNTY GUIDANCE CENTER
AND DEPARTMENT OF CRIMINAL JUSTICE ADMINISTRATION

Interrupted time-series design and quasi-experimental logic were employed to assess the behavioral effects of a prison disciplinary intervention on a population of approximately 2000 adult male inmates. The uncontrolled nature of the prison environment prevented experimental analysis of intervention effects. The quasi-experimental methodology applied in this study exemplifies how data, otherwise lost, can be retrieved. In addition, the process by which qualified conclusions can be drawn from the retrieved data is demonstrated.

A key philosophical and technical aspect of applied behavior analysis concerns empirical evaluation of treatment procedures. To this end, experimental designs utilizing randomization and control groups have been complemented by multiple baseline and reversal designs in the experimental analysis and evaluation of human endeavors (*cf.* Risley and Wolf, 1972). In all of these designs, one goal is to demonstrate convincingly that treatment procedures, not uncontrolled extraneous variables, are responsible for changes in dependent variables. In some cases, these experimental designs have not been employed when new treatment procedures are initiated. Thus, it becomes very difficult to isolate treatment effects from other sources of possible confounding influence. In these cases, one could reject any treatment data as being hopelessly contaminated by lack of experimental rigor. Alternatively, these contaminated data could be viewed from a quasi-experimental model.

Quasi-experimental designs are systematic attempts at evaluating plausible explanations for treatment effects (Campbell, 1969). Quasi-experimentation takes the posture that any experi-

ment is valid until proven invalid (Campbell, 1969). Threats to validity that invalidate an experiment are those that are as plausible as the hypothesized treatment. For example, if an experimenter wanted to analyze the effect of more liberalized liquor laws on the incidence of traffic accidents, he might examine the number of accidents after the change in light of the trend and variability in the number of traffic accidents before the law change. If there was a significant change in the dependent variable at the time of the law change, law change is a plausible explanation of the treatment effects. Alternatively, if there was a law enforcement crackdown on alcohol-related accidents at the same time as the law change, then the law enforcement crackdown also becomes a viable explanation of the change in number of traffic accidents.

The evaluation of treatment procedures in the quasi-experimental designs involves listing all known threats to validity of the interpretation of treatment effects. These threats are then evaluated in terms of the situation at hand and are either ruled out or accepted as a plausible explanation of the data. The result of this process is usually only a tentative conclusion concerning treatment effects, but at least a conclusion based on available data and knowledge of the experimental situation.

¹Reprints may be obtained from Dr. John F. Schnelle, Box 438, Middle Tennessee State University, Murfreesboro, Tennessee 37130.

The present paper employs a quasi-experimental design, the interrupted time series, in an attempt to evaluate retrospectively the behavioral effects of a prison policy change. The process employed in arriving at conclusions concerning the treatment effects demonstrates an experimental analysis of behavior model that is supplementary to traditional methods of such analysis.

The Tennessee State Prison at Nashville, Tennessee, initiated a policy change on July 15, 1969. This policy change specified that inmates who were behavioral problems would be sent to Brushy Mountain State Prison. The precise definition of behavioral problem was not specified. However, prison officials indicated that multiple offenders and inmates guilty of offenses resulting in personal injury would be the most likely candidates for transferral. Brushy Mountain State Prison was considered by the prison administrators to be a less attractive (*i.e.*, more aversive) facility than the Tennessee State Prison. This policy change was placed into effect over the entire prison on the same day. In such a situation, one of the only possible comparison bases is a record of prepolicy behavior with the postpolicy behavior.

Applications of Quasi-Experimental Designs

The interrupted time-series design was used to analyze the effects of the policy. This design extends a simple one-group "before" and "after" design to a series of observations at times removed from the immediate initiation of the policy, both before and after. In this situation, the behavior of the inmates was analyzed for each of seven different months before the policy change and for each of 23 different months after the change. This method is an alternative to evaluating data only immediately before and immediately after the policy change. Illustrations of such time-series analyses can be found in the literature. Campbell (1969), Campbell and Ross (1968), and Glass (1968), provide examples of the quasi-experimental philosophy and

the interrupted time-series design as applied to evaluation of a region-wide policy change concerning speeding violations. These studies analyzed the effect of a speeding crackdown in terms of pre- and posttreatment analysis within the population in which the crackdown took place and also by comparing the treatment population with nontreated control populations. Both within- and between-group comparisons were made with time-series data.

Lawler and Hackman (1969) employed a similar within- and between-population time series design in analyzing the effect of employee participation on the development of a pay incentive plan. Ross, Campbell, and Glass (1970) also employed a within-population time-series design to assess changes in automobile fatality and casualty rates after a government-initiated crackdown on driving under the influence of alcohol. In addition, different times of the week within the treated population (*i.e.*, commuting hours, weekend nights) were compared by time-series analysis in the Ross *et al.* study, the logic being that the crackdown on driving under the influence of alcohol should exert more of an effect on weekend nights rather than on commuting hours.

Parker (1963) combined aspects of all the preceding quasi-experimental designs in an attempt to assess the effect of television on public library circulation data. Parker compared communities without television to those with television, and he compared library circulation rates within a community before and after introduction of television. In addition, he took advantage of the fact that television was initiated at different times within all of the communities studied. Thus, he could also compare changes in library circulation rates within communities at a time when TV was first introduced and at a time when the influence was present but not changing. Also, he could compare between communities with one community having TV for a period of time and thus TV rates not changing, and the other community undergoing a TV growth period because of recent introduc-

tion of television. Parker called this an *ex-post facto* matched pair before and after design with switching replication. The design is an attempt to control for possible differences between communities that could result in the absence of randomized control groups, as well as for possible difference within communities over the different time periods of the experiment.

Gottman and McFall (1972) illustrated the application of a time-series analysis to the evaluation of the effects of two treatment procedures that were applied to two groups at the same time. These treatment procedures were later reversed and applied to the opposite treatment group. This is a case in which the time-series analysis was applied to complement a more conventional design.

It was not possible to apply any of the quasi-experimental control group procedures to the time-series analysis of the present experiment. Thus, it was necessary to confine analysis to a simple pre- and posttreatment difference within one population. Reasons for the absence of quasi-experimental control groups are made clear in the discussion section of the present paper.

METHOD

Subjects and Setting

Adult males incarcerated at the Tennessee State Prison for men in Nashville, Tennessee, for all or part of a period extending from January, 1969 to June, 1971 served as subjects. The average monthly population of Tennessee State Prison over the time period of this study was 1906. Virtually all lengths of sentences and classes of offenders were represented in the population. There was no systematic change in the prison population as to the type of offender during the time of the study, nor was there a seasonal variation in the type of inmates found in the prison population. The Tennessee State Prison is a 70-yr-old medium security prison. It is staffed by 295 guards, 41 health professionals, and 164 maintenance personnel.

Independent Variables

On July 28, 1969, the Tennessee State Prison initiated a new policy whereby all prisoners who were considered to be behavioral problems were transferred to Brushy Mountain State Prison, Petros, Tennessee. Brushy Mountain is a maximum security institution located in a remote area of East Tennessee. It is the impression of the prison administration and guards that the inmates regard Brushy Mountain as less attractive than the Tennessee State Prison. It is regarded as less attractive because of its geographic remoteness, fewer opportunities to have visitors, fewer opportunities for finding jobs, fewer training opportunities, and because of Brushy Mountain's reputation for maximum security. Because of the latter facts, and assuming the accuracy of the impressions of the prison administration and guards, transfer to Brushy Mountain may possibly be described as a timeout from positive reinforcement, and the policy change may then be viewed as a threat to use timeout procedure.

No administrative consideration was given as to how, or if, a prisoner could return to the Tennessee State Prison once he was transferred. What happened after transferral remained undefined. Furthermore, the policy change was not formally announced or posted. Instead, word was spread through the prison "grapevine".

Simultaneously with this transferral policy, the Tennessee State Prison discontinued the disciplinary technique of restricted diet. Restricted diet consisted of an unseasoned mixture of vegetables and meat. Men who were placed on restricted diet were simultaneously confined to an isolation cell for various periods of time. This solitary confinement procedure was not discontinued with the demise of the restricted diet technique. Solitary confinement remained as behavioral control procedure throughout all phases of the experiment. In addition, the disciplinary technique of reduction of good and honor time was maintained throughout the study.

This research is a retrospective attempt to describe the behavioral change among the prison

population that may have resulted from the adoption of the Brushy Mountain policy and the simultaneous discontinuance of the restricted diet policy. Possible behavioral effects of the Brushy Mountain policy are evaluated on the total prison population and also on the inmates who were not actually sent to Brushy Mountain. This latter population had been instructed as to the possibility of being sent to Brushy Mountain, and they had in the time period of the study observed men who were sent to Brushy Mountain.

Dependent Variables

Prison behavioral offenses were entered into a log book as they were observed by the prison guards. Punishable behavioral offenses are described in the Manual of Policy and Procedures, Department of Corrections, State of Tennessee.

Each offense was reviewed by a three-man disciplinary court which, in turn, passed judgement as to an appropriate punishment for each case. These decisions were entered into a log book. The average frequency of punishable behavioral offenses per month as recorded by the guards, type of offenses allegedly committed, number of transfers to Brushy Mountain, and the punishments given for these offenses served as dependent variables for the experiment.

Data Analysis

The behavioral records for 1969, 1970, and the first half of 1971 were analyzed by first graphing the total number of offenses for all prisoners during this 1969 to 1971 period. Secondly, these data were converted to mean numbers of daily offenses per man each month. This was done to control for changes in the total prison population, which could influence the number of monthly offenses. This data conversion was accomplished by dividing the total number of offenses for a month by the estimated total number of man-days spent at the Tennessee State Prison. The total number of man-days was estimated by taking weekly prison population counts for each month of the study. The average

of these weekly checks was taken as the monthly population figure. This monthly population figure was then multiplied by the number of days in the month to arrive at total monthly man-days. To illustrate this process, if the average of four weekly population checks in February was 1800, then 1800 would be multiplied by 28 to give the total number of man-days in the prison. The total number of man-days would then be divided into total recorded February offenses to give average monthly offenses per man.

Thirdly, the names of all men who were transferred to Brushy Mountain during the time period of the study were extracted from the transformed data described in the second stage. This extraction procedure was performed in order to make a valid comparison between the pretreatment data and the posttreatment data. To make such a comparison, it is necessary to equate the prison population in the 6.9-month interval before the policy was initiated and the 23.1-month interval after the policy. This equation could be accomplished only by removing from the records names of all men who were transferred out of the Tennessee State Prison to Brushy Mountain. This latter fact is true due to the fact that more men were transferred to Brushy Mountain in the posttreatment phase of the study than in a pretreatment phase. Presumably, the men transferred in the posttreatment phase were transferred specifically because they were behavioral problems. Men transferred in the first 6.9-month period were not transferred on the basis of any such behavioral criteria. Instead, the prison officials reported that men in the first 6.9-month period were transferred primarily because they were sentenced to long terms and because the population of the Tennessee State Prison was temporarily too large. Therefore, the populations in the pre- and posttreatment intervals could possibly be different and this difference would be due to the selection basis produced by the Brushy Mountain policy. The data collected from step 3 was graphed separately from the data collected in step 2 (see Figure 1). This separate graphing was

done to emphasize the fact that the behavioral data collected in step 3 are pertinent only to those offenders not sufficiently deviant to be transferred; that is, mild offenders. These data illustrate behavioral effects of the policy change on those people not actually punished by the policy of transfer to Brushy Mountain. Alternatively, the data collected from step 2 illustrate the behavioral frequency of all inmates in the prison. These step 2 data can be interpreted only in the light of two prominent intervening processes—the policy change and the selection out of the institution of those inmates who were sent to Brushy Mountain. This selection process would seem to ensure the fact that there would be a greater number of inmates more prone to committing behavioral offenses in the pretreatment phase than in the posttreatment phase of the study.

Fourthly, the total number of “one-time” behavioral offenders, “two-time” behavioral offenders, and “three-or-more-time” behavioral offenders was calculated separately for each of the five six-month intervals between January, 1969 and June, 1971 (see Figure 2). These latter three classes of offenders refer specifically to behavioral offenses committed inside the institution. Once again, names of men transferred to Brushy

Mountain were extracted from the step 4 data and plotted separately. This extraction was done for the same reason as enumerated in step 3.

Finally, the different types of offenses were classified into 11 basic classes that encompassed 98% of all types of offenses recorded in the institution. These classes are listed in Table 1. A frequency count of each of these offenses and of the different punishments for these offenses was made on a monthly basis. The term “different” punishments refers to the fact that only dissimilar punishments for a given behavior were counted. For example, if loss of honor time was the punishment seven times and isolation was the punishment five times for the offense of stealing, then the frequency of different punishment for stealing would be recorded as two. This “different” punishment statistic was employed in order to achieve an estimate of the consistency of punishments for the behavioral offenses.

RESULTS

Figure 1 illustrates the average number of daily offenses for all inmates not transferred to Brushy Mountain. These data are graphed over each month of the study. It is obvious that there

Fig. 1. Average number of offenses per inmate over a 30-month period. Brushy Mountain policy introduced July 28, 1969. Monthly data multiplied by a constant of 1000.

was a decrease between the months of July and August of 1969, which corresponded with the introduction of the Brushy Mountain Policy.

F ratios or *T* tests based on pre-intervention *versus* postintervention differences are not appropriate statistical tests for time-series data of the type illustrated in this study. Since most points in time-series data are not independent, the basic assumptions of either *F* or *T* ratios cannot be met. Maguire and Glass (1967) and Glass, Willson, and Gottman (*unpublished*) have developed statistical procedures to test for shifts in both level and slope of such interdependent time-series data. These statistical tests are based on models that take into account the dependency nature of the time-series data. Following the procedures outlined by Glass *et al.* (*unpublished*), the dependency nature of the time series was analyzed by means of autocorrelation of the original and first-order differenced pre-intervention data. The autocorrelation for the original data remained relatively large for several lags, suggesting some nonstationarity in this series. The autocorrelation of the first-order difference showed an abrupt drop to nonsignificance between the lag-one and lag-two autocorrelation, suggesting that the Figure 1 time-series data conform to an Integrated Moving Average (1, 1) model. Since there are so few pre-intervention points in the present study (7), a correlogram was also taken over both pre-intervention and postintervention data combined. The same type of correlational relationship depicted for the pretreatment data resulted from the second correlational analysis over the combined pretreatment and posttreatment data.

Employing a computer program designed by Glass *et al.* (*unpublished*), the maximum likelihood estimate of the dependency nature of the time series was found to be 0.6, with an estimated level effect of 1.923 and an estimated standard error level of 0.2967. Testing the hypothesis that the change from July to August of 1969 is significant yielded a *T* ratio of 2.81, which is significant at the 0.05 level. One concludes from this analysis that there was a signifi-

cant change in the level of offenses between July and August of 1969.

Evidence to confirm the fact that the present time-series data truly conform to an IMA (1, 1) model comes from the fact that the lag-one correlation of the first differenced data is expected to equal

$$-\frac{1 - Y}{1 + (1 - Y)^2},$$

with *Y* being the maximum likelihood value generated by the IMA (1, 1) analysis (Glass, 1968). In this case, the first-order differenced correlation was 0.39, and the value obtained by substituting *Y* in the latter formula is 0.34, a nonsignificant difference. It is concluded that the IMA (1, 1) model is a good fit of the time-series data in the present experiment.

The data presented in Figure 2 illustrate the frequency of three basic classes of offenders: "one-time" offenders, "two-time" offenders, and "three-or-more-time" offenders. These data are shown separately for each of the five six-month intervals analyzed and are based only on those inmates not transferred to Brushy Mountain. The percentage of the total number of offenses in each six-month block is listed for each class of offenders above the appropriate bar in Figure 2. It can be observed that the percentage of "two-time" and "three-or-more-time" behavioral offenders decreased after the Brushy Mountain policy was placed into effect in July to December, 1969. This decrease was maintained for a 30-month period despite an increase in the total number of offenses for the January to June period, 1971 (Figure 1). The large increase in total behavioral offenses for the latter period reflected a disproportionate increase in "one-time" offenders rather than "two-" and "three-or-more-time" offenders.

The average frequency of the 11 different classes of offenses is illustrated in Table 1. The frequency of both behavioral offenses and different punishments was analyzed on a monthly basis. No easily discernible trends were observed over the 30-month period of this study for

Fig. 2. Frequency of three classes of offenders shown for five six-month blocks. Figures above bars indicate percentage of each particular class of offender compared to other two classes within each block.

Table 1

Average frequency of 11 classes of behavioral offenders and average number of "different" punishments for each behavioral class.

OFFENSE CLASS	BEHAVIORS	OFFENSE FREQUENCY (averaged over a thirty month period)	DIFFERENT PUNISHMENT FREQUENCY
contraband major	possession of weapons, narcotics	6	37
contraband minor	possession of illegal money, literature, food, gambling devices	13	60
attacking guard	physical, verbal	2	12
sex	homosexual acts	2	18
intoxication	drugs, alcohol, glue	9	48
escape attempt	off of state property	4	2.2
stealing	possession of state or other inmates property	2	18
fighting	physical assault of other inmates	13	57
insubordination	refuse to follow orders	31	12.3
request lock up	inmate's own request	6	3.2
out of place	in prison area other than scheduled	13	59

either offense frequency or different punishment frequency. Thus, for ease of reporting, these frequency data were averaged over the entire 30-month period before entry into Table 1. It can be observed that there is little consistency in the punishments for any of the classes of behavior. There is an average of 4.4 different punishments for every 8.4 similar behaviors. This amounts to less than 50% consistency of punishment for any one behavioral class.

The inconsistency of punishments observed in Table 1 is also carried over to inconsistency in carrying out the Brushy Mountain (TO) policy. Data taken from the transfer records indicated that 10% of all "one-time" offenders, 32% of all "two-time" offenders, and 50% of all "three-or-more-time" offenders were sent to Brushy Mountain. Eight per cent of the men sent to Brushy Mountain had no behavioral offenses.

DISCUSSION

The significant decrease in the frequency of all behavioral offenses in July and August, 1969 coincided exactly with the initiation of the transfer to Brushy Mountain—restricted diet policy. The significance of this decrease and the fact that this was the largest decrease in offense frequency for the 30-month period of this study is evidence in support of the position that the policy produced a suppressant effect on "two-" and "three-time" offenders. The most plausible alternative explanations for the July to August decrease are taken from those suggested by Campbell (1969) and listed below:

1. History—the change may be due to events other than the experimental treatment that occurred at the same time or after treatment.
2. Maturation—the changes may be part of a long-term trend.
3. Instability—the July to August change may be no more than chance random variation.
4. Instrumentation—the change may be due to methods of measuring or recording the dependent variable.
5. Regression artifacts—if treatment was initiated when dependent varia-

bles were extremely high or low, then statistical probability is that it would be less extreme on subsequent measures, even though the intervening treatment is ineffectual.

6. Selection bias—the July to August difference is due to differential recruitment or selection of subjects for analysis.

1. In regard to the alternative explanation of history, it is obvious that termination of the restricted diet procedure simultaneously with initiation of the Brushy Mountain policy falls into this category. Such an obvious confounding cannot be defended except to say that it seems unlikely that the discontinuation of an aversive control procedure, restricted diet, either alone or in interaction with the introduction of a new behavioral control procedure, would produce such a drastic decrease. One cannot, however, completely discount the possibility that the termination of restricted diet had some influence on the July to August decrease. One must qualify any conclusions concerning the treatment effect by saying that the transfer to Brushy Mountain—restricted diet policy taken together may have produced the decrease. No other plausible explanation for the decrease can be found in the category of history.

2. Maturation seems not to be a feasible alternative explanation because there was no definite downward trend before the treatment. A more prolonged period of analysis before treatment would enable a more definite dismissal of maturation effects as a possible explanation of the time series effects. Unfortunately, pre-1969 data were not available.

3. The significant change in level calculated from the IMA (1, 1) analysis that was applied to the prison data is an effective argument against the instability hypotheses. Box and Tiao (1965) and Glass (1968) provided a complete discussion of this statistical model and its application to the problem of instability.

4. Instrumentation is perhaps the most plausible explanation of the present treatment results of all the hypotheses discussed. There were no direct controls in this situation over possible

changes in the reporting behavior of the guards that might have occurred after the policy change was initiated. In order to assess the relationship between the reporting behavior of the guards and the behavior of the inmates, it would be necessary to have independent reliability checks of the relationship between these two variables. Since this study was a retrospective analysis, no such reliability check was made. Furthermore, present attempts to make such an assessment are thwarted by the uncontrolled nature of the prison situation and the obtrusiveness that would result if an independent observer attempted to reliability-check the prison guards. Since no reliability check was made during the study, and since present reliability checks are prohibitive because of technical difficulties, it is necessary to qualify any conclusion concerning the treatment effect. Once again it becomes necessary to say that the July to August decrease was most plausibly produced by a combination of the Brushy Mountain—restricted diet policy and regression artifacts, and that the decrease could represent a decrease in the offensive behavior of the inmates or a reactive decrease in the reporting behavior of the guards.

Other data that possibly shed some light on the plausibility of such a reactive reporting change on the part of the guards can be found by analyzing the types of offenses reported both before and after the change. A listing of these classes of offenses is found in Table 1. In each case that an offense is reported by the guards, the reporting guard and the prisoner go before a reviewing board, which passes judgement on the case. In the example of fighting and sex offenses, all of the inmates involved must go to the board and report what happened. In the case of stealing, or possession of contraband major or minor, the objects that were stolen or illegally possessed must be produced or the case is dismissed. In the case of insubordination, the only evidence necessary to implicate the prisoner is the guard's personal testimony. The offenses reported most often under insubordination, are not following orders and disrespect for guards.

Thus, one might predict that if the guards reacted to policy change by reporting fewer offenses, then the behavior class that would show the greatest change would be insubordination. This is true because of the less obvious and more arbitrary nature of this offense class. Such was not the case in this instance, since all the offense classes showed a decrease after the policy change, and not just insubordination. Furthermore, frequency in none of the offense classes decreased significantly, thus demonstrating that the total offense decrease was spread evenly over all offense classes. The latter statistical tests were accomplished by applying the integrated moving average statistic to each offense class individually. Thus, if reporting behavior of the guards changed, it changed uniformly over both obvious and unobvious offenses. The preceding fact is still not sufficient to eliminate re-active recorder changes as a plausible explanation of the results. The offense class data suggest only that perhaps recording behavior did not change. Reliability checks are still needed.

5. Regression artifacts are not easily dismissed as possible influencing agents in these time-series data. In this situation, the treatment was initiated on July 28, a month that corresponded with the second highest offense frequency in the pretreatment phase. Statistically, a decrease is expected in the next measuring period whether or not a treatment is initiated. In principle, the degree of this expected regression can be computed from the autocorrelation of the series. In this case, however, there are not enough data to do this with confidence. In spite of the fact that regression cannot be discounted, it is still unlikely that the entire July to August decrease can be accounted for by regression artifacts. The largest previous decrease was between January and February, 1969, a total drop of 47 offenses. The July to August decrease was a drop of 67 offenses.

6. The selection of inmates out of the population who are most prone to committing behavioral offenses is the last validity threat to the in-

terpretation of the present time-series data. This threat implies that the July to August decrease can be accounted for by the fact that those inmates most prone to committing behavioral offenses, as well as repeated offenses, were progressively removed from the Tennessee State Prison population. Thus, the pre-intervention data would have a higher proportion of severe behavioral problems and repeated offenders than would the postintervention data.

The data transformation procedure described in step 3 of the method section was an attempt to control for this selection hypothesis. This transformation procedure produced time-series data only on those inmates who were never transferred out of the Tennessee State Prison. Thus, the preintervention and postintervention time series data are comparable and not contaminated by a selection process that progressively removed the most troublesome inmates from the population. It must be noted that even though the July to August decrease cannot be explained by selection factors, the present data are only representative of the behavior of the inmates who were not considered severe enough behavior problems to be transferred. Thus, these data can only be generalized to those inmates who are not categorized the worst behavioral problems in the institution.

Thus far, the most plausible explanation for the July to August decrease is the initiation of the Brushy Mountain—restricted diet policy, regression artifacts, and possible changes in the recording behavior of the guards. Data that bear on the relative plausibility of the latter three explanations can be found by measuring the frequency of the occurrence of transfer to Brushy Mountain. If the transfer policy is a factor in explaining the July to August decrease, then one would expect that a significant change would take place in Brushy Mountain transfers on the month that the effect was first measured. The statistic of relevance in this case would be the percentage of Brushy Mountain transfers compared to other types of penalties. This is true because the raw frequency of punishment would

change depending on the number of offenses committed, and hence provide a spurious estimate of changes in punishment policies in the prison.

It was the case in the present situation that there was a large increase in the percentage of Brushy Mountain transfers in August. The increase was from a percentage level of 3% to a level of 33%.

Explanations of the data following the July to August decrease are even less clear than the July to August change. The February to March 1970 increase, which took the offense frequency back to near prepolicy baseline levels, and the April to May 1970 increase, in which offense frequency exceeded baseline levels, are the most dramatic points subsequent to the July to August decrease.

Perhaps the best hypothesis for the 1970 and 1971 increases can be drawn from the punishment consistency data in regard to timeout. Ten per cent of all "one-time" offenders, 32% of all "two-time" offenders, and 50% of all "three-or-more-time" offenders were sent to Brushy Mountain. These contingencies are particularly meaningful in view of the fact that the chance of being sent to Brushy Mountain as a "one-time" offender ($p = 0.10$) was not dramatically different from the fact that 8% of the men transferred to Brushy Mountain had zero behavioral offenses. One would not expect that Brushy Mountain would have permanent behavior-suppressing properties when administered in such a contingency system.

The temporary seven-month suppression that did occur might be attributed to the fact that at the beginning of the timeout policy, it was uncertain as to exactly what contingencies were to be employed. Interviews with disciplinary personnel revealed that "troublemakers" were to be disciplined by timeout. The behavioral referents of "troublemakers" remained undefined. It might, therefore, be predicted that behavioral offenses would show a suppression in that period when the inmates and prison administration were not certain of the contingen-

cies between timeout and behavior. When the contingencies were demonstrated in practice to be weak and thus the uncertainty removed, the behavior increased back to baseline levels.

If the latter argument is correct, one would expect that in this experiment the increase back to baseline should be accounted for mostly by an increase in "one-time" offenders. This is true, since the punishment probability for this group proved to be insignificantly different from the punishment probability for zero behavioral offenses. Data to support this conception are illustrated in Figure 2. It is obvious from these data that the first dramatic increase in total behavioral offenses (Figure 1, February to March, 1970) can be accounted for almost entirely by a corresponding increase in "one-time" offenders (Figure 2, January to June, 1970). Both total monthly frequencies and "one-time" offender frequencies were suppressed below baseline for the first six months of the experiment, when the inmates were perhaps not certain as to the actual contingencies, and both were elevated above baseline levels thereafter.

The percentage of "two-" and "three-or-more-time" offenders, unlike "one-time" offenders, showed a permanent decrease throughout the experiment. This could indicate that the transfer to Brushy Mountain and the contingency between multiple offenders and the probability of transfer to Brushy Mountain were effectively suppressing multiple offense behavior. Taken together with the information concerning "one-time" offenders, the data suggest that transfer to Brushy Mountain might be an aversive stimulus only when delivered according to the contingency system that existed for multiple behavioral offenses. It is not an aversive stimulus when delivered according to the contingencies that existed for "one-time" offenders; *i.e.*, it does not have aversive properties when such consequences are not in fact applied.

In summary, the July to August decrease cannot be logically attributed solely to initiation of a behavioral policy in that time period. The conclusion that the policy produced a change is very

plausible, but there are also three other plausible explanations. The explanations fall under the category of instrumentation, history, and regression artifacts. Experimental conditions necessary to separate out these effects in future applications of such prison policies as the Brushy Mountain transfer policy would be simple to fulfill in a more planned, systematic fashion.

The best method to analyze the treatment in question—the Brushy Mountain transfer and restricted diet policy—would be to administer the treatment to one cell block at a time in multiple baseline (*cf.* Baer, Wolf, and Risley, 1968) fashion with different cell blocks serving as the different baselines. Time-series methodology and statistical procedures could supplement this multiple baseline across subjects design. (Risley and Wolf, 1972).

Regression artifacts could be alleviated by random selection of times to administer the treatment to cell blocks. Reliability checks on the guards' reporting behavior would be feasible due to the smaller number of people and more limited areas of environment involved at any one time in the multiple baseline design.

This multiple baseline design seems to be superior to picking another prison population to serve as a control group. This is true because of obvious differences between prison populations in Tennessee and because reliability checks of guard reporting behavior are difficult to accomplish even in one prison. In addition, the prepolicy differences between cell blocks within a prison in Tennessee are most certainly smaller than differences that would exist between prisons.

Unfortunately, no solid conclusions can be drawn from the data in the present study. Errors in the implementation of this prison policy change becloud evaluation of the change. Methods of initiating the policy change that would allow logical evaluation are suggested by the errors described in this study.

The contribution of the study lies in the description of the quasi-experimental model of evaluation that was applied to these data. This model prevents unwarranted conclusions about

the data, defines limits of data interpretation, and also suggests methods by which definite conclusions can be made. These quasi-experimental designs are strong supplements to the more commonly employed analysis of behavior techniques. Behavioral scientists in general and correctional psychologists in particular should be cognizant of such important analysis tools.

REFERENCES

- Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimension of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 20-34.
- Box, G. E. P. and Tiao, G. C. A change in level of a non-stationary time-series. *Biometrika*, 1965, 52, 181-192.
- Campbell, D. T. Reforms as experiments. *American Psychologist*, 1969, 24, 409-429.
- Campbell, D. T. and Ross, H. L. The Connecticut crackdown on speeding: time series data in quasi-experimental analysis. *Law and Society Review*, 1968, 3, 33-53.
- Glass, G. V. Analysis of data on the Connecticut speeding crackdown as a time series quasi-experiment. *Law and Society Review*, 1968, 3, 55-76.
- Glass, G. V., Willson, V. L., and Gottman, J. M. *Design and analysis of time series experiment*. Unpublished manuscript, Laboratory of Educational Research, University of Colorado, 1972.
- Gottman, J. M. and McFall, R. M. Self-monitoring effects in a program for potential high school dropouts: a time series analysis. *Journal of Consulting and Clinical Psychology*, 1972, 39, 273-281.
- Lawler, E. E. and Hackman, J. R. Impact of employee participation in the development of pay incentive plans: a field experiment. *Journal of Applied Psychology*, 1969, 53, 467-471.
- Maguire, T. O. and Glass, G. V. A program for the analysis of certain time series quasi-experiments. *Educational and Psychological Measurement*, 1967, 27, 743-750.
- Parker, E. B. The effects of television on public library circulation. *Public Opinion Quarterly*, 1963, 27, 578-589.
- Risley, T. R. and Wolf, M. M. Strategies for analyzing behavioral change over time. In J. Nesselroade and H. Reese (Eds.), *Life-span developmental psychology: methodological issues*. New York: Academic Press, 1972.
- Ross, H. L., Campbell, D. T., and Glass, G. V. Determining the effects of a legal reform: the British "breathalyzer" crackdown of 1967. *American Behavioral Scientist*, 1970, 13, 493-509.

Received 27 June 1973.

(Revision requested 16 January 1974.)

(Final acceptance 1 May 1974.)