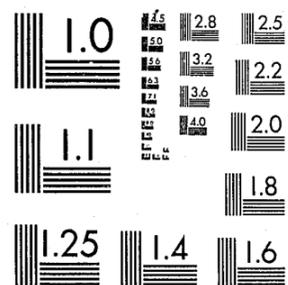


National Criminal Justice Reference Service



This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice  
United States Department of Justice  
Washington, D. C. 20531

DATE FILMED

10/08/81

78159

MF-1

MEASURING CORRECTIONS PERFORMANCE

EXECUTIVE SUMMARY

THE OSPREY COMPANY  
Raleigh, North Carolina

1980

U.S. Department of Justice  
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by  
Public Domain  
U.S. Department of Justice  
to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

Prepared under Grant Number 78-NI-AX-0130 from the National Institute of Justice, U.S. Department of Justice. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

SCOPE AND APPROACH

This report summarizes the approach to performance measurement proposed in the book, Measuring Corrections Performance. The recommended approach recognizes the diversity of adult corrections programs and the environments within which those programs exist as well as the multiplicity of performance measurement users and uses. It seeks to help people interested in measuring corrections performance define their measurement needs and develop a corrections performance measurement system appropriate to those needs.

The framework for measuring corrections performance presented in this report resulted from: (1) studying what correctional agencies do and who the major actors are; (2) analyzing problems and issues confronting measurement; and (3) constructing and interpreting measures for the five major program areas--jails, prisons, probation, parole, and community-based programs. The results of the research in each of these areas are summarized below.

RESULTS OF RESEARCH

I. What Corrections Agencies Do and Who the Major Actors Are

Several factors best describe corrections agencies:

- . the goals of corrections
- . the relationship of corrections to the criminal justice system
- . the organizational alignments of correctional agencies
- . the roles of persons in corrections
- . the specific description of the primary functions, the type of clientele, and the process of the agency

Detailed information on these factors is available in the book, Measuring Corrections Performance (Chapter II). For the purpose of this report, it is sufficient to note that it is this information which provides the groundwork for the performance measurement framework described in the next two sections, below.

II. What the Measurement Problems and Issues Are

A. SETTING THE SCOPE AND FOCUSING THE PERFORMANCE MEASUREMENT SYSTEM

Performance measurement is a broad, nebulous concept that needs to be defined and structured before one can measure performance. A person developing a performance measurement system for adult corrections programs should resolve several issues before thinking about specific measures. The issues discussed in this section first concern the role that goals, theories, and the ability to control program outcomes should play in shaping the performance

65181

measurement system. Second, they concern the role that the performance measurement system itself should play in influencing program performance.

Goals: What effect will goals have on performance measurement system design?

Goals may be defined as broad, general statements of desired conditions external to programs that provide the basic purposes for which programs were authorized and funded. If performance measurement were to be based upon a rational model of decision-making, the first step in developing a performance measurement system would be identifying the goals against which performance is to be compared. Though this step seems easy, there are several questions that need to be considered before the performance measurement system is built around a set of goals.

First among these questions is, "Whose goals should be recognized?" The potential users of performance information include the public, legislators, chief executives, agency heads and administrators, program managers, planners, budgeters, employees, and clients. These groups, if asked to agree upon a single set of goals for a corrections program, would probably be unable to do so. The public, for example, might be primarily interested in the program's ability to incapacitate and punish offenders and make the community a safer place in which to live, while the offender might be primarily interested in the quality of the services that the program makes available to him.

One may think of corrections goals in terms of broad outcomes, such as revenge or retribution, restraint, reform, or rehabilitation, reintegration into society, and restitution.<sup>1</sup> Goals of individuals or groups interested in corrections programs, however, may be unrelated to any of these broad outcomes. A community might support a prison because it absorbs a large part of that community's work force. Community groups might feel that an important goal of the prison is to provide employment to community residents. Private businesses in the community might look to the prison as a source of revenue through sales of food, medical and dental supplies, maintenance supplies, materials for prison industries and through providing contract services. Business groups, then, might believe that an important goal of the prison is to provide business opportunities to the community.<sup>2</sup>

Within the organization one may be confronted with three types of goals. First are the official, stated goals, which in their broadest form might be stated in terms such as these: to rehabilitate offenders, to reduce subsequent criminal activity, to punish the guilty, to provide restitution to the victims of crime. Next, there may be management goals that make possible attainment of the official, stated goals. At their broadest level, management goals might be stated in terms such as these: to secure the resources necessary to support the organization's programs adequately, to

<sup>1</sup> See Carter, McGee, and Nelson (1975: 12-13) for one such discussion of correctional goals.

<sup>2</sup> Perrow (1978) contains an excellent discussion of different goals ascribed to organizations.

build and maintain employee morale, and to maintain internal stability within the organization. Third, individual employees and clients may have their own goals, such as to have a pleasant place to work, to advance one's career, to build up one's retirement fund, or to "do easy time." All these goals may affect the organization's performance.

If all these types of external and internal goals affect the performance of corrections programs, should progress toward all these goals be monitored through the performance measurement system? One approach to deciding the scope of the performance measurement system might be to limit those goals used as guides in identifying what is to be measured to corrections-oriented goals (e.g., retribution, rehabilitation, restitution) and to exclude non-corrections-oriented goals (e.g., employment, business opportunities, career advancement, doing "easy time"). This approach is broad enough to include information addressing the following sorts of questions asked about corrections programs: What did the program spend? What did the program produce? How was the product produced? How good was the product? What was the cost per unit of product? What was the cost per unit of benefit? What needs remain unmet? The advantage of such a broad approach of performance measurement is that it includes the information felt important by many of the potential users, such as funding agencies, program managers, chief executives, legislators, and the public.

Although such a broad approach to developing a performance measurement system is conceptually appealing, such a system is likely to be expensive to implement. It would be more economical to design a system that responds to the specific informational needs of selected users. In practice, the performance dimensions included in the system may depend upon who pays for its implementation and how much the payor is willing to spend. Such a practical resolution of the scope problem has the disadvantage of leaving some groups of people interested in corrections performance with performance data that do not fit the decisions they must make. For example, performance measurements designed to answer the questions raised by the program manager may not be relevant to the decisions the legislator must make.

However the question of whose goals are to be recognized is resolved, there is likely to remain the problem of what to do when goals are inconsistent with each other. Our approach would be to recognize that corrections is one of many policy areas that reflect inconsistent and sometimes conflicting values held by our society. It is not the task of performance measurement (or of designers of performance measurement systems) to resolve these conflicts. Such conflict resolution is a function of the political process. Performance measurement can best serve that process by identifying multiple outcomes of correctional programs and leaving the assessment of their relative importance to those people who will use performance information.

Theories: What effect will theory have on performance measurement system design?

A theory is "an integrated body of propositions, the derivation of which leads to explanation of some social phenomenon" (Denzin, 1970: 5). Theories are important when deciding what to measure for three reasons:

1. Theories shape the content of programs.
2. Theories influence our expectations of outcomes.
3. Theories influence our interpretation of the meaning of the performance measurements obtained.

Because of the influence that theory has upon what is to be measured and how measurements are to be interpreted, one must be aware of the effect that holding a particular theory is likely to have upon the content of a corrections performance measurement system. Where there is no consensus about which theories are correct, as there is not for most corrections programs, performance measurements can be considered from multiple theoretical perspectives.

Which theories about corrections programs and their assumed effects should be taken into account when designing a performance measurement system? Many theories in the social science literature are relevant to corrections programs. The researcher is likely to want to focus upon that subset of theories upon which his own research is based. The practitioner may have developed and implemented his program upon some explicit theory contained in the social science literature. Or the practitioner may have his own theory of action, theory of practice, or theory in use.<sup>3</sup>

Control: Should the system measure only those outcomes that corrections agencies can control?

Should performance measures be developed only for those events over which actors in corrections agencies can exert total or near-total control? This question is explored in two steps. First, should performance measurement be restricted to program dimensions over which a single agency or actor has total control? Second, should performance measurement be restricted to program outcomes over which corrections programs have total control?

In the United States today, there are few corrections activities that a single actor or agency controls exclusively. Several governmental agencies share responsibility for funding and managing most programs. It is concluded, therefore, that restricting performance measurement to program dimensions over which a single agency has control would so restrict the scope of performance measurement that the information produced would be trivial compared to the questions being asked about program performance.

<sup>3</sup>These three terms are used as defined by Argyris and Schon (1974: 6, 11). "A theory of action is a theory of deliberate human behavior," which states what a correctional program ought to do to achieve certain results. A theory of practice "consists of a set of interrelated theories of action that specify for the situations of the practice the actions that will, under the relevant assumptions, yield intended consequences." A theory in use is a theory of deliberate human behavior inferred by the way the practitioner behaves. A practitioner's theory in use can be different from his espoused theory of action.

An alternative approach would be to focus upon what a single program can control rather than what a single agency or actor can control. Jointly, the various actors that influence the resources, processes, outputs, and outcome objectives for a single program can control that program's direct outputs. The program's performance can be measured in terms of these outputs, even though the responsibility for this program belongs to no single actor.

Restricting performance measurement to those events over which a single program has total or near-total control, however, excludes almost all program outcomes or impacts. Program outcomes, such as an offender's post-release criminal activity, are affected by environmental factors beyond the control of the program. Yet failing to measure program outcomes means that a program's varied constituent groups will not know how the public is better off as a result of corrections programs.

Another approach should be considered. In this approach, program activities would be treated as contingent conditions preceding outcomes. Further, because corrections programs are contingent conditions influencing outcomes, outcomes are an appropriate dimension for describing corrections program performance. One who adopts the third approach must confront the problem of how to sort out the impact of a corrections program upon an outcome--say post-release criminal activity--from all the other factors that affect that outcome. Otherwise, one might inappropriately interpret outcome measurements to infer program success or failure. We suggest multivariate statistical analysis as the most practical method of separating program impact from other influences upon outcomes.

#### Measurement Affects Performance

Performance measurement is not a neutral managerial tool. Management control systems, for example, include performance measures for the explicit purpose of detecting deviations from plans or standards so that, when program processes malfunction, managers can take action to bring operations back on course. Neither should it come as any surprise that measures designed to compare performance to goals focus an organization's effort upon those activities that are measured.

The researcher cannot design a performance measurement system that can aid policymaking without also affecting performance. One should be sensitive to the effect that performance measurement has upon staff behavior. Including measures that foster activity at the expense of program results should be avoided. If a performance measure cannot be a neutral tool, one might at least try limiting measures to ones that affect behavior positively.

#### B. DEVELOPING PERFORMANCE MEASURES

This section concentrates on how a designer of a performance measurement system would use our approach to select, define and assess performance measures. First, we discuss how performance measures are to address specific questions; second, we outline how performance measures relate to particular program concepts; and, finally, we suggest how one can decide which performance measures to include in a performance measurement system.

## Deciding What to Measure

In deciding what to measure, the first step is identifying the questions that people want answered about a program's performance. The questions about performance most frequently raised address efficiency, cost-effectiveness, equity, service quality, unmet need, and conformance with government policies.<sup>4</sup> Which of these performance dimensions should be built into a corrections performance measurement system? Because data collection is expensive, cost may encourage one to restrict the scope of performance measurement to a subset of these performance dimensions. Before doing so, the designer should carefully consider who will use the performance information in order to collect information most beneficial to that entity.

Various constituents can use performance information as ammunition in the political process to support or attack corrections programs. Once performance information is collected, it is hard to limit the public's access to it or to control the way that information is used in the political process. Consequently, the designer should consider who is likely to want performance data and the purposes for which they are likely to use it.

Natural constituents for performance information about corrections programs include researchers, planners, budgeters, public interest groups, legislators, funding agencies, and chief executives, as well as corrections agency heads and program managers. If the designer limits the performance dimensions measured to those of greatest interest to a couple of these groups the information provided will probably not adequately answer some of the questions other groups ask about program performance.

What would be the consequences of not answering these other performance questions? Actors in the political process will not withdraw from the process because they do not have performance information. They will either proceed to maneuver without performance information or will use or misuse whatever performance information exists. Neither will researchers stop doing research because they lack performance information. The designer should keep in mind that limiting the performance dimensions included in the measurement system will probably result in some measurements being used (misused?) to answer other questions than those the measurements were designed to answer.

## Identifying Performance Concepts

Before identifying the specific information needed to answer the selected performance questions, the designer would summarize the program concepts that relate to the question being addressed. Flowcharting or diagramming is a convenient method for relating program processes to outcomes, making assumptions about cause-effect relationships explicit and identifying the performance concepts that are important to measure.<sup>5</sup>

<sup>4</sup>Definitions of each of the performance measure and performance comparison terms used are included in Appendix C of the book, Measuring Corrections Performance.

<sup>5</sup>Examples of flowcharts or diagrams used to identify concepts are found in the book, Measuring Corrections Performance (Chapter III).

As previously suggested, the designer may want to consider performance measurement from multiple theoretical perspectives. Performance measurement budgets, however, may be insufficient to permit measuring program performance from multiple perspectives. The designer should at least make clear the concepts he feels are important to be measured and the cause-effect assumptions that relate these concepts to the corrections program whose performance is being measured.

## Assessing the Adequacy of Potential Performance Measures

After deciding what to measure and identifying the performance concepts, the designer would think about ways of measuring each concept, research the corrections evaluation and measurement literature, and pull together a list of potential measures.

Deciding which of these potential measures to accept and which to reject may be aided by applying a uniform set of criteria to evaluate each measure. These criteria would define the premises upon which measures are compared in order to establish their relative desirability. The most appropriate criteria will vary, depending upon how one intends to use measures.

Criteria frequently suggested for rating potential performance measures fall into four categories.<sup>6</sup> Criteria for technical adequacy relate the potential measure to the concept it measures and permit assessing the measure in terms of how valid, reliable, and accurate the measurements are likely to be. Practicality criteria address concerns about the cost and ease of obtaining data. Two other categories consider utility from a general perspective and from the perspective of the specific use intended for the measure. Knowing how comparable, sensitive, and clear the measure is can give one an idea of the range of programs and constituents for which a measure might be useful. Timeliness and relevance of performance measurements to decisions, on the other hand, can be judged only within the context of specific uses.

Potential measures need to be rated by people who understand the situation in which performance measurements will be used. The rater can design a rating strategy for identifying measures that meet the constraints of his particular situation.

One should keep in mind that this rating process is basically subjective. The ratings strategy gives a rater a systematic way of thinking about factors that render a potential measure satisfactory or unsatisfactory. One can use a scoring scheme to produce a single numerical rating for comparing measures that relate to the same concept and selecting measures that rate higher than some predesignated cutoff. If this procedure seems too mechanical, one can simply apply the criteria to obtain insights about the measures' strengths and weaknesses without producing total scores.

If none of these strategies seems desirable, a more unstructured approach could be used. One could, for example, first sort measures into "suitable"

<sup>6</sup>Figure III-7, Criteria for Rating Potential Performance Measures, can be found in the book, Measuring Corrections Performance. For a survey of literature on criteria, see Grizzle (1979a).

and "unsuitable" categories and then summarize the factors that led to the judgment that some measures were suitable and others not.

### III. How Measurements Are Interpreted and Measures Are Constructed

This section summarizes factors that affect how one interprets measurements, examines bases for comparing measurements, suggests various ways of constructing performance measures, and presents a way of using statistical models to estimate program effects.

#### A. INTERPRETING MEASUREMENTS

Lists of performance measurements are not by themselves of much value. Before using these measurements, one must decide what they mean. This decision requires that one interpret the measurement within the context of additional information. We have already emphasized the important role that theory plays in shaping the meaning of performance measurements. When a program's assumed cause-effect relationships are explicit and when performance measures relate to specific concepts within the cause-effect framework, the direction in which a measurement should change to be interpreted as an improvement in performance should be clear.

In addition to theory, several other matters need considering when interpreting measurements. These factors include timing, self-correcting cases, learning curves, and participant dropouts. Timing is especially important when measuring outcomes that lag behind program operations. The learning curve phenomenon may also affect program performance. If program operations are such that one can expect improved performance to result from experience, one should consider the program's developmental stage when interpreting measurements. Program dropouts also need to be considered when interpreting program outcomes. A high dropout rate, if ignored, can lead to judging program effects only in terms of that portion of participants who were most successful.

In addition to the special care in interpretation posed by the timing of measurements, self-correcting cases, learning curves, and program dropouts, attributing outcomes to a specific corrections program rather than to other factors generally poses a problem of interpretation. A later section of this report illustrates the multivariate statistical models that we believe present the most feasible approach to dealing with this attribution problem.

#### B. USING MEASUREMENTS TO JUDGE PERFORMANCE

Measurements describe performance but do not by themselves evaluate it. To judge how well a program is doing, performance measurements must be compared with other information. This information may take the form of standards, goals or objectives, optimal or technically efficient performance levels, or the performance of other programs.

When comparing performance to standards, one would conclude that performance at or exceeding the level prescribed in the standard is satisfactory. Performance measurements at levels below the standard would indicate need for improvement. Similar conclusions could be reached, using quantified goals or objectives instead of standards.

The other bases for comparison, technical efficiency and interprogram comparisons, require more detailed discussion than do standards or goals.

### Technical Efficiency

Technical efficiency means producing the maximum output from a given input bundle. This concept can be applied to corrections programs to estimate the reduction in cost possible if technical efficiency prevailed. But before technical efficiency concepts are applied to correctional programs, several questions need to be answered.

The first question that needs to be raised is, "Is the finding of the most technically efficient program useful to anyone?" To assume either that it is "good" *per se* to be technically efficient or that the technically inefficient programs ought to emulate the technically efficient programs requires that we agree on two points. The first point of agreement is that the inputs, as measured, adequately capture the important aspects of the process. A second point upon which we must agree is that the quantity of a particular output adequately captures the output of the entire process.

It may be argued that the quantity of a particular output of a process does not capture important qualitative variations in the outputs of different programs in the process. This argument is especially relevant when the audience for the research is concerned both with technical efficiency and with allocative efficiency (whether the marginal benefit is equal to the marginal cost and output is produced at the lowest cost). If this concern is to be addressed, then it will probably be necessary to enrich the production function by including a vector of output quality attributes.

It may also be argued that the measures for the input bundle do not capture important process differences between programs. The variables may define quantitative combinations of the inputs but not describe *how* the inputs are combined. The quality of the output can be affected by the way resources are used, not simply the quantity and proportions of the inputs used.

### Interprogram Comparisons

Performance measurements most usefully indicate how well a program is performing when measurements can be compared with each other. Interprogram performance comparisons are most appropriate when these conditions are present:

1. When process measurements are used to compare performance, programs should share common processes.
2. When efficiency or product measurements are used, programs should share common products.
3. When quality measurements are used, programs should share common service characteristics.
4. When equity measurements are used, potential client groups should be similar.

5. When effectiveness or cost-effectiveness measurements are used, the types of outcomes expected should be similar among programs compared to each other.
6. Programs should use the same definitions, data collection and reduction procedures, and measurement display formats.
7. Data collection and reduction techniques should be practical and relatively cheap.
8. Programs must have an opportunity to explain unusual situations.
9. Accurate and timely data collection and reporting should be rewarded.<sup>7</sup>

#### Aggregating Multiple Outcomes

When programs have more than one outcome in terms of which performance can be compared, assessments can be made in two general ways. In the first way, the outcomes are simply arrayed and the user must decide how much importance to attach to each outcome when judging program performance. In the second way, weights are attached to each outcome and these weighted outcomes are summed to provide a single performance measurement. Several researchers have presented methods of determining and applying these weights to various performance dimensions (Keeney and Raiffa, 1976; Edwards, Guttentag, and Snapper, 1975; Rohrbaugh and Wehr, 1978; Rohrbaugh and Quinn, 1979). The technique presented in the book, Measuring Corrections Performance, is that developed by Edwards (Edwards and Guttentag, 1975; Edwards, Guttentag, and Snapper, 1975).

Taken in the method's simplest form, the steps are to identify program outcomes, determine the relative importance of each outcome, estimate the extent to which each program attains each outcome, multiply each outcome attained by its relative importance, and sum these products in order to calculate the utility for each program.

#### C. CONSTRUCTING MEASURES

Measures may be constructed as simple counts; ratios; percentages, or unit costs; indices; or models that estimate a measure as a function of several other variables.

##### Simple Counts

Simple counts are frequently used to measure cost, amounts of work done, quantity of product, and outcome.

##### Ratios, Rates, and Percentages

By taking two simple counts and dividing one by the other, one can construct ratios, percentages, and unit costs.

<sup>7</sup>Points 6 through 9 were adapted from Dressel (1976: 92).

#### Unit Cost

Unit cost measures can be used to make both efficiency and cost-effectiveness comparisons.

#### Indices

Indices that aggregate simple counts, percentages, or ratios are another way of constructing performance measures. Rather than developing unit costs for each activity, one can use an index to construct a single product measure that includes several activities.

Deciding the relative importance of the different activities is the challenge in constructing such an index. If each activity is weighted equally, a distortion of effort may be presented. This distortion of effort could be avoided by weighting each activity by the average time required to complete that activity. Incentives for providing quality service can also be introduced by adding measures of service characteristics to the index.

Multiple outcome measures can also be aggregated to provide a single index. As with the activity index, the most difficult problem in constructing an outcome index is deciding the relative importance of the individual performance measures. The decision theoretic techniques mentioned in the section on aggregating multiple outcomes are one way of deriving the weights.

#### Models

We have illustrated how performance measures can be constructed as simple counts, ratios, percentages, unit costs, and indices. None of these methods has dealt with the problems noted several times in this report: How can one attribute a change in one of the performance measurements to a particular program when other factors also affect the phenomena being measured? We have suggested that the most practical method of approaching this problem is through multivariate statistical modeling. The next section explains how one can use models to isolate program effects.

#### D. USING MODELS TO ISOLATE PROGRAM EFFECTS

The proper use of multivariate statistical control requires two things: (1) models of the performance measure of interest, which indicate for what it is necessary to control; and (2) appropriate statistical techniques which allow control for the factors identified.

##### A Method of Developing Models of Performance Measures

Adequate models of corrections performance measures will tend to be eclectic in nature, drawing insights from the work of researchers and practitioners in a large number of areas. In order to develop eclectic models of corrections performance measures, the researcher or practitioner must have a set of criteria with which to judge the adequacy of different models. One set of criteria which we have found valuable in evaluating models of some performance measures includes: (1) completeness, (2) universality, (3) transferability, (4) explanatory powers, (5) data availability, and (6) understandability.

Note that we combine traditional criteria for assessing theoretical adequacy with practical concerns about data availability and understandability. We believe that this combination is essential when developing models for performance measures. Only models which are theoretically reasonable and can be both estimated and communicated will allow those who wish to compare and contrast the performance of corrections organizations to do so.

To illustrate the way in which models for corrections performance might be developed, we develop models for (a) the extent and timing of post-release criminal activity and (b) the wages of individuals who have participated in a corrections program.

#### Development of a Model for the Extent and Timing of Post-Release Criminal Activity

Because theoretical models of criminal behavior are subject to waves of acceptance and rejection and are based on the perspective of a single discipline, they are limited in their individual usefulness for developing models of corrections performance measures.

No single model found during our survey of the literature attains consistently high rating for completeness, explanatory power, universality, transferability, data availability, and understandability. We feel that this situation is mainly due to the fact that the theories surveyed tended to be developed within a given discipline. Further, different models are designed to explain different types of crime.

Many of the variables suggested as being important by the theories may be measured empirically. Using the variables empirically found to be associated most strongly and consistently with criminal behavior<sup>8</sup> and combining this insight with the insight gathered from theories, the following model for the extent and timing of post-release criminal behavior was developed.

Extent and timing of criminal activity = f (family, perhaps measured by marital status or changes therein; job and residential stability; family values and activities; criminal record; mental health commitments; socioeconomic status, perhaps measured by occupation, wages, and educational attainment; employment stability, as a measure of work satisfaction; opiate or alcohol abuse; age; sex; race; IQ; age at first arrest; type of release; type and quality of correctional programs; length of time served before release; effectiveness of the criminal justice system; genetic and physiological factors; and the environment in which the individual currently finds her or himself)

<sup>8</sup>Table IV-2 in the book, Measuring Corrections Performance, lists "Factors Predictive of Future Criminal Conduct." This list was greatly influenced by the survey work of Service (1972); Gillespie (1975), Blumstein, Cohen and Nagin (1978) and Monahan (1980a).

#### Development of Model for Wages

Currently, there are three major schools of thought on wage determination: (1) neoclassical economic, (2) human capital, and (3) institutional or structural. No single model attains consistently high ratings for completeness, universality, transferability, explanatory power, data availability, and understandability. As was the case for our model of criminal behavior, we develop an eclectic model of wages, combining the insights from the theories of wage determination and the empirical work survey.<sup>9</sup> We arrive at the following model for the wage rate of individuals released from corrections programs:

Wage = f (industry of employment, occupation of employment, geographic area where employed, rate of unionization for job, permanence of job, opportunity for movement on job; skill development possibilities of job, minimum wage, race, sex, criminal records, age, education, previous work experience, marital status, number of dependents, availability of other income, physical condition, mental condition, motivation, level of addictive problems, correctional experience)

#### A Method of Selecting Statistical Techniques to Estimate Models for Correctional Performance Measures

When selecting an appropriate technique for statistically estimating models of corrections performance measures, one should carefully consider three factors: (1) the nature of the model (causal vs. exploratory), (2) the distribution of the dependent variable, and (3) the nature of the explanatory variables. Given these factors, one often finds a number of statistical techniques which are good candidates for use in estimation. One set of criteria which we have found valuable in evaluating potential estimation techniques for models of some corrections performance measures includes: (1) technical appropriateness, (2) methodological strength, (3) flexibility, (4) sensitivity, (5) the availability of significance tests, (6) transferability, (7) costs, and (8) understandability.

To illustrate the way in which estimation techniques for performance models might be selected, we will select statistical techniques to estimate the models of the extent and timing of criminal activity, and wages which we developed in the previous section.

#### Selection of a Statistical Technique to Estimate Our Model for the Extent and Timing of Post-Release Criminal Activity

The most commonly used measure of timing is the length of time until an offense occurs. This measure, when combined in an appropriate manner for a group of correctional releasees, also provides a measure of the extent (proportion returning to criminal activity) of criminal activity. This variable requires considerable care in statistical analysis as it is nonnegative, skewed and truncated from above. The nonnegativity arises from the fact that it is not

<sup>9</sup>Table V-7 in the book, Measuring Corrections Performance, lists the empirical results of previous studies of wage and income.

possible to observe negative times until recidivism. The distribution of this variable is generally quite skewed as those who return to crime generally do so quite quickly, although lower rates of failure occur throughout a follow-up period. The truncation of the variable arises because we cannot observe a value of the dependent variable greater than the length of time for which an individual's activities are followed.

A number of authors have suggested alternative techniques for analyzing this variable.<sup>10</sup> It is possible to evaluate these different statistical techniques using the set of evaluative criteria mentioned above. According to our assessment, the truncated lognormal technique scores most highly on appropriateness, methodological strength, flexibility, and the availability of tests of statistical significance. However, the simpler Stollmack-Harris, Maltz-McCleary, and Witte-Schmidt OLS techniques score more highly on understandability. The ultimate choice of a method thus must rest on the relative importance of understandability and more technical statistical concerns.

#### Selection of a Statistical Technique to Estimate Our Model of Wages

Like the timing of criminal activity, the wages of an individual require considerable care in statistical analysis. This dependent variable is non-negative and truncated at zero. Further, it is quite likely that there will be a substantial "pile-up" of observations at zero (the wage an individual receives if he is unemployed). In a working paper (Bass, 1979), we evaluated methods assuming a truncated normal distribution developed by Tobin (1958) and Amemiya (1973) (a Tobit model), ordinary least squares, and a two-stage procedure developed by Heckman (1976 and 1979). Using the set of evaluative criteria mentioned above, it is apparent that the Tobit technique developed by Tobin and Amemiya scored highest on technical grounds, while ordinary least squares analysis scored highest in terms of cost and understandability. If sufficiently trained personnel and computer facilities are available, we recommend that Tobit analysis be used to estimate wage models.

When sufficiently trained personnel and adequate computer facilities are not available, we recommend that correctional releasees' wages be recorded until most if not all releasees are employed (our experience (Witte, 1975) indicates that more than 98% of prison releasees find jobs within two months of release). The proposed model for wages, augmented by the addition of a variable indicating length of time until first job, could then be estimated using ordinary least squares analysis (OLS). Tobin's work (1958) indicates that the biases introduced by using OLS will be greater the nearer the values of the dependent variable approach the truncation point, zero in the case of wages. Heckman's work (1976, 1979) indicates that the greater the probability that a zero wage rate is observed the greater will be the bias involved in using OLS. This insight led us to suggest that releasees be followed until most if not all had been employed. Heckman also shows that using OLS will lead to estimates of standard errors which are too small. Hence, when the wages model is estimated using OLS, one should utilize stringent tests of statistical significance (e.g.,  $\alpha = .01$  or  $\alpha = .001$ ).

<sup>10</sup>Stollmack and Harris (1974), Maltz and McCleary (1977), Bloom (1978), Witte and Schmidt (1977).

#### FUTURE RESEARCH NEEDS

The agenda for future research calls for the development of an operative performance measurement system for adult corrections programs. While this approach encompasses jail, prison, probation, parole, and community-based programs, the research proposed below can focus on some subset of these programs if one's interests or resource constraints so dictate. The first set of research questions pursues some of the thorny theoretical issues that surfaced during the research effort summarized in this report. The second set of research issues deals with statistical models for generating information about the efficiency and impact of corrections programs. The last set of research topics deals with issues that one must face when implementing a performance measurement system.

##### I. Deciding What to Measure and How to Interpret Measurements

The disagreement over what is most important to measure and how one interprets measurements will continue to frustrate performance measurement research unless the problem is confronted directly. We suggest that three researchable questions be pursued.

##### What Are the Critical Operations in Corrections Programs upon Which Performance Measures Ought to Focus?

Corrections operations should be examined within the framework of an explicit theory about what correctional agencies ought to do and how they ought to go about doing it. Using the theories relevant to the particular correctional agency being examined and the outcomes of that agency, as well as the prevailing philosophies of the programs, a number of variables that affect both what that agency does and how measurements of its operations would be interpreted to judge performance can be abstracted. From this list of variables, one could select the ones that seem most important as summary descriptors of program differences pertinent to performance measurement research. How measurements of operations would be interpreted to judge agency performance might depend upon the combination of these variables as it applies to a given agency. This theoretical framework should guide identification of critical operations upon which the correctional agency's performance measures ought to focus.

Flowcharts can be used to relate activities to each other by showing the patterns through which work flows from one task or activity to another. Next, the researcher can estimate the amount or level of work for each operation and the rate of flow from one operation to another. Having this information permits identifying those critical operations where changes in capacity or rate of flow might have substantial impacts on agency performance. It is for these critical operations that measures could be devised that would allow an agency manager to diagnose operational problems that would hinder agency performance.

##### What Measures Can Different Constituent Groups Agree upon as Being Adequate Measures of Performance?

Answering this question requires that the researcher (a) generate a set of potential measures that members of constituent groups can consider and (b) use some method for assessing the adequacy of these potential performance measures.

It is problematic whether a single set of measures can be developed that will be acceptable to all constituent groups. We suggest looking at patterns of agreement and disagreement among three groups--researchers, criminal justice practitioners, and funding agency staff.

#### How Does the Relative Importance of Different Performance Dimensions Vary among Constituent Groups and over Time?

Performance dimensions that relate directly to corrections operations include cost, quantity of product, quality of service, efficiency, and equity of distribution. In the long run, however, researchers should not ignore effectiveness dimensions. As an aid to ranking future research efforts, the relative importance of these performance dimensions to different constituent groups should be researched.

Two techniques, multiattribute utility theory and social judgment analysis have been used to develop relative values, or weights, for multiple objectives. With the multiattribute utility theory approach, judgments about the relative importance of different performance dimensions would be elicited separately from the scoring function. Using social judgment analysis, one derives the relative weights of performance dimensions and the scoring function simultaneously.

The researcher could elicit weights from a group of people using both methods. By entering for each rater the weights obtained from the multiattribute procedure as coefficients in the regression equations for the hypothetical profiles, a second set of overall performance ratings can be obtained. If the correlations are high between these two sets of ratings, the cheaper procedure would be justified. If the correlations are low, the researcher should try to determine the source of error and select the procedure believed to be most valid.

This same decision theoretic approach can be used to research two other questions relating to specific performance dimensions. The first question is, what equity standard should be used when comparing the distribution of services? The second question relates to multiple outcomes. What is the relative importance of different outcomes, such as increased employment, reduced criminal activity, and increased family stability?

## II. ESTIMATING EFFICIENCY AND IMPACTS

In addition to researching critical operations in corrections programs, identifying measures that different constituent groups can agree upon, and learning how the relative importance of different performance dimensions varies, we recommend that further research be conducted on two performance dimensions--efficiency and impact. The two sections that follow propose further research applying statistical models to efficiency and impact measurements.

### Average Cost and Frontier Cost Models

Further research applying traditional average cost and frontier cost functions can be used to (a) identify which local units within a state system are operating most efficiently, (b) diagnose the factors associated with low cost per offender, (c) estimate the reduction in cost that would be possible

if technical efficiency prevailed, and (d) predict future costs, given estimated changes in offender population, offender characteristics, quality of corrections program processes, and so on.

Research on corrections cost functions requires developing a statistical model for corrections programs relating the average cost of production to the level of direct output or product, the quality of direct output, the service conditions under which production takes place, and the cost of inputs (e.g., facilities and manpower) to the production process.

### Multivariate Statistical Outcome Models

While researchers have recently given considerable attention to measuring outcomes when evaluating corrections programs, outcome measurements are still not generally available. Even when outcome measurements are available, one usually does not know to what extent the corrections program (rather than other factors) contributed to the outcomes. We recommend proceeding with three types of outcome-measurement research:

- 1) Develop theoretical and empirical models for outcomes in addition to the labor market and post-release criminal activity outcomes.
- 2) Build upon the labor market and criminal activity models by developing simultaneous equation models.
- 3) Collect the data required to estimate these models.

## III. USING PERFORMANCE MEASUREMENTS

Before corrections agencies implement performance measurement systems, they need to know how to collect the data on a regular basis, what the measurement system would cost, and what incentives are necessary for corrections actors to collect the data and use the performance measurements.

### Understanding Incentives for Measuring Performance

Generally, researchers need to identify and classify the different factors that serve as incentives or disincentives for corrections actors to develop and use performance information. Any jurisdiction seriously contemplating performance measurement would be well advised to inventory the likely incentives existing in its corrections agency. Building incentives into the performance measurement effort is as important to the success of the effort as is the technical work.

### Developing and Testing Procedures for Regular Data Collection

Little research has been completed that tests data collection procedures for process, equity, efficiency, and cost-effectiveness measures. We would recommend that researchers devote more effort to this area as soon as we learn which measures different constituent groups agree are adequate for these performance dimensions.

### Developing Model Performance Measurement Systems

This last phase of research links the researcher to the practitioner. The researcher would work with a few corrections agencies in order to see how interests, measurement, and utilization capabilities are likely to vary among agencies; assess the level of interest in implementing measurement systems and identify obstacles to developing them; develop model or sample performance measurement systems tailored to the needs of those agencies worked with; and prepare descriptions of these model performance systems that the National Institute of Justice could distribute to interested corrections agencies.

### SUMMARY AND CONCLUSIONS

Performance measures for a corrections program should be developed within a conceptual framework. This framework should identify:

- . the correctional program whose performance will be measured,
- . that program's stage of development and the types of performance information appropriate to that stage,
- . who is asking what questions about the program's performance and how they expect to use answers to these questions,
- . who will pay for performance measurement and what restrictions will the payor place on the scope and content of the performance measurement effort,
- . which (and whose) information needs will the performance measurement system serve,
- . what will be the likely consequences of not serving some information needs,
- . to what benchmarks performance will be compared to judge performance,
- . what the program does and how it goes about doing whatever it does,
- . what theory guides one's choice of what to measure and how to interpret measurements, and
- . what specific concepts need measuring.

Developing measures without such a framework is likely to produce data that potential users perceive as useless or that are subject to misinterpretation. We conclude that no single list of performance measures is appropriate to all adult corrections programs and that performance measures can best be developed within frameworks tailored to specific programs.

Potential performance measures need to be assessed against some set

of criteria to select the strongest measures for data collection. The most important criteria will vary, depending upon such factors as how the measures will be used and the amount of money available for collecting data. We suggest that criteria likely to be important are validity, reliability, accuracy, cost and ease of data collection, comparability, sensitivity, clarity, relevance to decision, and timeliness.

When interpreting performance measurements, the user should keep in mind the theoretical concepts that the performance measures represent. One should also take into account factors that distort measurements, such as program dropout rates, learning curves, self-correcting cases, and when measurements are made. Measurements simply describe performance, but comparisons permit evaluations of performance. Comparisons can be made to standards, goals, objectives, targets, other programs, or to measurement of the same program made at earlier times.

Performance measures can be constructed as simple counts, ratios, percentages, or unit costs. Indices are ways of aggregating several measures into a single overall measure. When dealing with process, product, service characteristic, and distribution measures, one should take special care when constructing the index to avoid unintentionally producing an index that distorts program effort. Managers, by the measures they include in these indices and the relative weights they give to them, can provide employees incentives to emphasize particular activities and/or service characteristics and to serve offenders with certain characteristics (e.g., those having greatest need).

One can also combine several outcome measures into an index. Again, one should take special care when developing weights for the outcome measures to ensure that they represent the relative importance that performance measurement users ascribe to them. When different users do not agree upon the relative importance of the outcomes being combined into an index, two or more indices may be required. In this event, each index would include the same measures but have a separate set of weights attached to those outcomes.

While correctional actors may have substantial control over program operations, programs are usually only one of many factors that influence the changes in offenders toward which correctional programs aim. One must attribute some portion of changes in outcomes, such as future criminal activity or economic productivity of ex-offenders, to specific correctional programs in order to estimate the impact that correctional programs have upon these outcomes. We suggest statistical control through multivariate modeling as the most practical way of estimating these impacts.

We suggest one way in which models of correctional performance measures may be developed. Specifically, we suggest that different models for correctional performance measures be evaluated on the following criteria: (1) completeness, (2) universality, (3) transferability, (4) explanatory power, (5) data availability, and (6) understandability. If no single model clearly dominates on the basis of these criteria, we suggest that eclectic models of the performance measure be developed. We illustrate a method of developing such eclectic models by developing models for the timing and extent of post-release criminal activity and post-release wages.

We next suggest a method for selecting a technique to estimate models for correctional performance measures. Specifically, we suggest that different

statistical techniques for estimating models of correctional performance measures be evaluated on the basis of the following criteria: (1) technical appropriateness, (2) methodological strength, (3) flexibility, (4) sensitivity, (5) the availability of significance tests, (6) transferability, (7) costs, and (8) understandability. If no one statistical technique clearly dominates others on all criteria, we suggest that the individual researcher or practitioner decide upon the relative importance of the various criteria and select the technique which (s)he feels ranks highest on the most important criteria. We illustrate the use of our criteria for selecting statistical techniques by selecting statistical techniques for estimating the models we developed for the extent and timing of criminal activity and wages. Specifically, for our model of the extent and timing of criminal activity, we suggest that the truncated lognormal technique be used for estimating if technical criteria are most important and that ordinary least squares analysis be used if transferability, costs, and understandability are more important. For our model of post-release wages, we suggest Tobit analysis if technical concerns dominate and ordinary least squares analysis if transferability, costs, and understandability dominate.

We have concluded this report on measuring correctional performance by proposing an agenda of research that one could undertake within the theoretical framework developed here. This proposed research is premised upon the belief that one should answer these questions before deciding to implement a performance measurement system:

- . What are the critical operations in correctional programs upon which performance measurements ought to focus?
- . What measures can different constituent groups agree upon as being adequate measures of performance?
- . How does the relative importance of different performance dimensions vary among constituent groups over time?
- . What incentives and disincentives exist for people to collect data for measuring performance and to use performance information? How can one build additional incentives into an organization and reduce existing disincentives?

Other researchers are currently testing data collection procedures for various performance measures. Once the issues listed above have been researched, it will be possible to integrate the results of their research on data collection procedures with that outlined in this report. Additional research will then be needed for testing data collection procedures for some performance dimensions not now receiving much attention. These dimensions are equity of service distribution, process, efficiency, and cost-effectiveness.

## BIBLIOGRAPHY

Amemiya, T. "Regression Analysis When the Dependent Variable Is Truncated Normal." Econometrica, 41 (November 1973), 997-1016.

\_\_\_\_\_ and M. Boskin. "Regression Analysis When the Dependent Variable is Truncated Lognormal with an Application to the Determinants of the Duration of Welfare Dependency." International Economic Review, 485-496.

Argyris, Chris and Donald A. Schon. Theory in Practice: Increasing Professional Effectiveness. San Francisco: Jossey-Bass, 1974.

Bass, J. "Statistical Methods for Estimating Models of Labor Market Performance." Raleigh, N.C.: The Osprey Company, 1979.

Blair, Louis H., et al. Monitoring the Impacts of Prison and Parole Services: An Initial Examination. Washington, D.C.: The Urban Institute, 1977.

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

Blumstein, A., J. Cohen, and D. Nagin (eds.) Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates. Washington, D.C.: National Academy of Sciences, 1978.

Cain, Glen. "Regression and Selection Models to Improve Nonexperimental Comparisons" in Carl A. Bennett and Arthur A. Lumsdaine (eds.) Evaluation and Experiment: Some Critical Issues in Assessing Social Programs. New York: Academic Press, 1975.

Carter, Robert M., Richard A. McGee, and E. Kim Nelson. Corrections in America. Philadelphia: J.B. Lippincott, 1975.

Colorado Department of Corrections. Work Papers for 1979-80 Budget Request, 1979.

Denzin, Norman K., ed. Sociological Methods: A Sourcebook. Chicago: Aldine, 1970.

Dressel, Paul L. Handbook of Academic Evaluation. San Francisco: Jossey-Bass, 1976.

Edwards, Ward and Marcia Guttentag. "Experiments and Evaluation: A Re-examination" in Carl A. Bennett and Arthur A. Lumsdaine (eds.) Evaluation and Experiment: Some Critical Issues in Assessing Social Programs. New York: Academic Press, 1975.

\_\_\_\_\_ and Kurt Snapper. "A Decision-Theoretic Approach to Evaluation Research" in Elmer L. Struening and Marcia Guttentag (eds.) Handbook of Evaluation Research, vol 1. Beverly Hills: Sage, 1975.

Gillespie, R.W. Economic Factors in Crime and Delinquency, final report submitted to the National Institute for Law Enforcement and Criminal Justice, 1975.

Grizzle, Gloria A. "Criteria for Rating Corrections Performance Measures." Raleigh, N.C.: The Osprey Company, 1979a.

\_\_\_\_\_. "A Typology of Performance Measures for Corrections Programs." Raleigh, N.C.: The Osprey Company, 1979b.

Heckman, J. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and Simple Estimator for Such Models." The Annals of Economic and Social Measurements, 5 (January 1979), 475-492.

Keeney, Ralph L. and Howard Raiffa. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York: John Wiley, 1976.

Maltz, M.D. and R. McCleary. "The Mathematics of Behavioral Change." Evaluation Quarterly, 1(1977), 421-438.

Monahan, J. The Clinical Prediction of Violent Behavior. Washington, D.C.: U.S. Government Printing Office, 1980, forthcoming.

Perrow, Charles. "Demystifying Organizations." in Rosemary C. Sarri and Yeheskel Hasenfeld (eds.) The Management of Human Services. New York: Columbia, 1978, 105-120.

Rohrbaugh, John and Robert Quinn. "Evaluating the Performance of Public Organizations: A Method of Developing a Single Index." Unpublished Manuscript. Albany, N.Y., 1979.

\_\_\_\_\_ and Paul Wehr. "Judgment Analysis in Policy Formation: A New Method for Improving Public Participation." Public Opinion Quarterly, 42:4 (1978), 521-532.

Service, P. The Recidivism of Persons Released from Facilities of the North Carolina Department of Corrections during January-July, 1968. Raleigh, N.C.: North Carolina Department of Corrections, 1972.

Stollmack, S. and C.M. Harris. "Failure Rate Analysis Applied to Recidivism Data." Operations Research, 22(1974), 1192-1205.

Sykes, Gresham. The Society of Captives: A Study of A Maximum Security Prison. Princeton, N.J.: Princeton University Press, 1958.

Tobin, J. "Estimation of Relationships for Limited Dependent Variables." Econometrica, 26(January 1958), 24-36.

Witte, Ann D. Work Release in North Carolina: An Evaluation of its Post-Release Effects. Chapel Hill, N.C.: Institute for Research in Social Science, 1975.

\_\_\_\_\_ and P. Schmidt. "An Analysis of Recidivism, Using the Truncated Lognormal Distribution." Journal of the Royal Statistical Society, Series C (Applied Statistics), 26(1977), 302-311.

END