Multijurisdictional Sentencing Guidelines Program Test Design

National Institute of Law Enforcement and Criminal Justice
Law Enforcement Assistance Administration
U.S. Department of Justice

Program Test Design documents are developed by design groups composed of representatives of the National Institute of Law Enforcement and Criminal Justice and various other LEAA and Department of Justice program offices. The documents are prepared with contractual assistance, and are reviewed by a panel of experts conversant with the critical research and operational issues in the topic area.

The design group for the Multijurisdictional Sentencing Guidelines Program Test Design consisted of the following individuals:

National Institute of Law Enforcement and Criminal Justice, LEAA

Anthony Pasciuto, Chairman
Carolyn Burstein
Joel Garner
G. Martin Lively
Jay Merrill
Benjamin Shapiro

Consultant support was provided by Joseph Calpin, Michael Fischel, Joseph Sasfy, and Lawrence Siegel; The MITRE Corporation, McLean, Virginia.

Advisory Committee members were:

Alfred Blumstein, Professor, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, Pennsylvania
Don Gottfredson, Dean, School of Criminal Justice, Rutgers University, Newark, New Jersey
Jack McCarthy, Director of the Sentencing Disparity Research Project, Administrative Office of the Court, Trenton, New Jersey
George H. Revelle, Judge, Superior Court of the State of Washington for the County of King, Seattle, Washington
Leslie T. Wilkins, Professor, School of Criminal Justice, State University of New York at Albany, Albany, New York

December 1978

Henry S. Dogin Acting Administrator Law Enforcement Assistance Administration	Blair G. Ewing Acting Director National Institute of Law Enforcement and Criminal Justice

PREFACE

The National Institute of Law Enforcement and Criminal Justice, the research arm of the Law Enforcement Assistance Administration, is sponsoring a field test of the concept of sentencing guidelines which has been proposed as a means of:

- reducing disparity in sentencing; and
- articulating judicial sentencing policy.

The basis for the field test is a Program Test Design, a document with detailed specification of selected program elements. The goals of each field test effort are to determine the effectiveness of these elements or program strategies in multiple settings and to examine their transferability to other jurisdictions.

A number of single, local court systems of general jurisdiction have experimented with the development and implementation of sentencing guidelines. In order to assess the feasibility of this concept beyond a single jurisdiction, the National Institute has devised a multi-jurisdictional test of sentencing guidelines which will involve selected courts within two or three states. In each state sentencing guidelines will be developed and implemented in from three to five local courts of general jurisdiction. Both processes of development and implementation as well as their outcomes will be evaluated by the Institute. There are three primary objectives of the field test:

- to evaluate the effectiveness of sentencing guidelines as a mechanism for enhancing sentencing consistency both within and across different jurisdictions within the same state;
- to test the feasibility of developing and implementing sentencing guidelines in a multijurisdictional setting; and
- to provide a body of knowledge for jurisdictions looking for a means to structure judicial decision-making.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

I. INTRODUCTION

A. Sentencing Guidelines: A Response to Disparity

Disparity, or unwarranted variation, in the sentencing of convicted offenders has long been observed, debated and explored, but until recently little has been done to provide judges with a structured method to reduce or eliminate the problem. Throughout the United States today, there is an increasing awareness of the need for greater equity in sentencing, i.e., that similar offenders committing similar offenses receive similar sentences.[1] To reduce unwarranted sentencing variation a variety of reform approaches have been suggested: these include such judicial initiatives as appellate review of sentencing and sentencing councils, as well as legislative reforms such as "flattime," mandatory and presumptive sentencing. However, these approaches have been criticized for encroaching on judicial sentencing prerogatives, lacking flexibility or being too cumbersome and costly.

Sentencing guidelines appear to present a viable alternative means to the courts for structuring judicial decisionmaking and reducing unwarranted sentencing variation. The use of sentencing guidelines has two primary goals. They are:

- to increase equity in sentencing, i.e., to reduce unwarranted variation (disparity) while retaining judicial discretion to individualize sentences; and

- to articulate an explicit sentencing policy and to provide a regular basis for policy review and change.

Other goals include the provision of an information tool for new or rotating judges and the promotion of increased visibility and understanding of the sentencing process to those outside the judiciary.

---

[1] See among others, New York State Special Commission on Attica, Attica: The Official Report of the New York State Special Commission on Attica, (New York: Bantam Books, 1972); David Fogel, We Are the Living Proof, The Justice Model for Corrections, (Cincinnati: W. H. Anderson, 1975); Marvin Frankel, Criminal Sentences: Law Without Order, (New York: Hill and Wang, 1972); The Twentieth Century Fund Task Force on Criminal Sentencing, Fair and Certain Punishment, (New York: McGraw-Hill, 1975); The American Friends Service Committee, Struggle for Justice: A Report on Crime and Punishment in America, (New York: Hill and Wang, 1971); and Andrew von Hirsch, Doing Justice: The Choice of Punishment, (New York: Hill and Wang, 1975).

B. The Conceptual Basis of Sentencing Guidelines

The sentencing guidelines concept is best described as a collaborative effort by the judiciary and by researchers to develop an empirically based system to reduce unwarranted variation and to thereby increase equity in sentencing. On the basis of a statistical analysis of past sentencing decisions, it establishes a yardstick for comparing offenders and the offenses they have committed in order to determine if they are similar in terms of key characteristics. A two-dimensional grid, with offender and offense dimensions, then supplies a recommended sentence for cases assessed as similar. Figure 1 depicts a hypothetical sentencing grid for a statutory class of felony crimes.

It is important to note that a sentence derived from the guidelines is not mandatory. Judges retain the discretionary authority to deviate from the guideline sentence and to impose individualized sentences. In those cases where a sentence falls outside that sentence range suggested by the guidelines, this system requires that the judge provide a specific written reason for doing so. According to the guideline concept, the periodic analysis of these reasons provides a basis for adapting the guidelines to a changing environment. This periodic analysis can assist judges in determining what changes might be necessary in the guidelines to reflect, accommodate and document shifts in court policy.

Sentencing guidelines represent an extension of a guideline methodology originally used to structure the decisionmaking process of the U.S. Parole Board.[2] In the context of this methodology, guidelines are defined as:

> a system of data which functions as a tool in assisting decisionmakers in arriving at individual and policy determinations. It accomplishes this purpose by using some form of equation(s) to summarize the link among the main concerns, or focal dimensions, of decisionmakers and their decisions.[3]

---

[2] Don M. Gottfredson, Leslie T. Wilkins, and Peter B. Hoffman, Guidelines for Parole and Sentencing: A Policy Control Method, (Lexington, Massachusetts: Lexington Books, 1978), pp. 13-41.

[3] Leslie T. Wilkins, Jack M. Kress, Don M. Gottfredson, Joseph C. Calpin, and Arthur M. Gelman, Sentencing Guidelines: Structuring Judicial Discretion, (Washington, DC: February 1978), p. 4.

FIGURE 1. SENTENCING GUIDELINES GRID EXAMPLES**

EXAMPLE A

Offense Score	0-1	2-3	4-5	6-7	8-9	10-11
4	4-6 yrs.	5-7 yrs.	6-8 yrs.	7-9 yrs	8-10 yrs.	8-10 yrs.
3	OUT	OUT*	3-5 yrs.	4-6 yrs.	5-7 yrs.	6-8 yrs.
2	OUT	OUT	OUT*	2-4 yrs.	3-5 yrs.	4-6 yrs.
1	OUT	OUT	OUT*	OUT*	1-3 yrs.	2-4 yrs.

Offender Score

EXAMPLE B

Offense Score	0-1	2	3	4	5	6+
5	24-36 mo.	36-48 mo.	48-60 mo.	60-72 mo.	72-84 mo.	84-96 mo.
4	16-24 mo.	24-36 mo.	36-48 mo.	48-60 mo.	60-72 mo.	72-84 mo.
3	12-16 mo.	16-24 mo.	24-36 mo.	36-48 mo.	48-60 mo.	60-72 mo.
2	OUT*	12-16 mo.	16-24 mo.	24-36 mo.	36-48 mo.	48-60 mo.
1	OUT	OUT*	12-16 mo.	16-24 mo.	24-36 mo.	36-48 mo.

Offender Score

EXAMPLE C

Offense Score	1	2	3	4	5	6	7	8	9	10	11	12+
6	1 yr.	1.5 yrs.	2.0 yrs.	2.5 yrs.	3.0 yrs.	3.5 yrs.	4.0 yrs.	4.5 yrs.	5.0 yrs.	5.5 yrs.	6.0 yrs.	7.0 yrs.
5	OUT*	1 yrs.	1.5 yrs.	1.75 yrs.	2.0 yrs.	2.5 yrs.	3.0 yrs.	3.25 yrs.	3.5 yrs.	4.0 yrs.	4.5 yrs.	5.0 yrs.
4	OUT	OUT	OUT	OUT*	1 yrs.	1.5 yrs.	2.0 yrs.	2.25 yrs.	2.5 yrs.	2.75 yrs.	3.0 yrs.	4.0 yrs.
3	OUT	OUT	OUT	OUT	OUT	OUT*	1 yrs.	1.5 yrs.	1.75 yrs.	2.25 yrs.	2.75 yrs.	3.5 yrs.
2	OUT	OUT	OUT	OUT	OUT	OUT	OUT*	1 yr.	1.5 yrs.	2.0 yrs.	2.5 yrs.	3.0 yrs.
1	OUT	OUT	OUT	OUT	OUT	OUT	OUT	OUT*	1 yr.	1.5 yrs.	2.0 yrs.	3.0 yrs.

Offender Score

*The offender is a potential candidate for an alternative sentence.

**See Gelman, et al for other examples of sentencing grids.

Guidelines as defined above have their genesis in the assumptions of decisionmaking theory. First, guidelines recognize that most decisions are based on limited information. The selection and processing of information beyond a certain point appears usually not to affect a particular decision. Too much data may render a decision so complex that the result is an inefficient use of information. Thus, sentencing guidelines typically use only the key factors in sentencing to define the offense and offender scores.

Second, when repeated decisions are involved, guidelines assume two levels of decisionmaking: the individual level, on which decisions are made one at a time; and the policy level, which represents an aggregation of individual decisions. Thus, while sentencing guidelines are used to structure individual decisions, they are, in fact, a reflection of the aggregate analysis of decisions at the policy level and it is assumed that:

> at the policy level it is possible to derive an equation to predict decisions on the basis of case information. This predictive ability may be interpreted as a description of latent or implicit policy which in turn provides the basis for the articulation of that policy.[4]

Third, statistical techniques provide a means of determining those factors which most influence decisions at both the individual and policy levels. Such techniques allow the synthesis and analysis of large quantitites of information from various sources. Thus, the descriptive capabilities of statistical methods provide a good basis for the identification of factors that influence sentencing and, therefore, for the development of guidelines. Although guidelines are based on the statistical identification of policy, the concept also incorporates the need for decisionmakers to make decisions on an individual basis because of unique factors.

C. Sentencing Guidelines Research

Recognizing the potential importance of sentencing guidelines, the National Institute of Law Enforcement and Criminal Justice (NILECJ) in 1974 initiated a three-year project to study the feasibility of

---

[4] Joseph C. Calpin, Jack M. Kress, Marilyn A. Chandler, Mona Margarita, Susan Mitchell-Herzfeld, Arthur M. Gelman, and Barbara A. Broderick, The Analytical Basis For the Formulation of Sentencing Policy, (Unpublished, January 1978), p. 3.

the concept.[5] The major question addressed by this project was whether a system of sentencing guidelines could be developed that was acceptable to judges in a single jurisdiction.

To test this feasibility, NILECJ funded the development and implementation of sentencing guidelines in the criminal courts of four individual jurisdictions: the Denver District Court (Denver, Colorado); the Essex County and Superior Courts (Newark, New Jersey); the Circuit Court of Cook County (Chicago, Illinois); and Maricopa County Superior Court (Phoenix, Arizona).[6] In each of these sites a research team from the Criminal Justice Research Center (CJRC) in Albany, New York, worked with an advisory group of the judiciary to develop a specific sentencing guideline model. Each model was based on an empirical analysis of past sentencing decisions and employed a set of two-dimensional grids.

Analysis of the experiences encountered in the development and implementation of sentencing guidelines has led to the conclusions that:

- simple methods could be developed which would provide a workable decisionmaking aid to a judge at sentencing;
- judges were willing to cooperate in the development of sentencing guidelines; and

---

[5] The feasibility stage of the Sentencing Guidelines Project was supported by Grant No. 74-NI-99-0054 awarded to the Criminal Justice Research Center, Inc., Albany, New York by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice; the test implementation stage by Grant No. 76-NI-99-0102. For further information concerning these developmental efforts see: Leslie T. Wilkins, Jack M. Kress, Don M. Gottfredson, Joseph C. Calpin, and Arthur M. Gelman, Sentencing Guidelines: Structuring Judicial Discretion, (Washington, D.C.: February 1978); Joseph C. Calpin, Jack M. Kress, Marilyn A. Chandler, Mona Margarita, Susan Mitchell-Herzfeld, Arthur M. Gelman, and Barbara A. Broderick, The Analytical Basis for the Formulation of Sentencing Policy, (Unpublished, January 1978); and Arthur A. Gelman, Jack M. Kress and Joseph C. Calpin, Establishing a Sentencing Guidelines System: A Methods Manual, (Washington, D.C.: November 1977). These documents are the basis for this test design and are essential references for any guideline effort. All of these documents are available from the National Criminal Justice Reference Service (Law Enforcement Assistance Administration).

[6] Using the same basic methodology developed for these four jurisdictions, the City of Philadelphia and the State of New Jersey are also developing sentencing guidelines with LEAA support.

- judges were willing to actually use guidelines in sentencing offenders.[7]

D. The Test Design

Based on the previous efforts, it appears that sentencing guidelines are feasible at least within a single local jurisdiction.[8] However, because guidelines are intended to reduce unwarranted variation (i.e., disparity), their fullest application would be at a statewide level, since criminal laws are established by state legislatures with the intent of establishing consistent sentencing practices throughout the state. Further, much of the criticism regarding disparity has been and continues to be directed at differences in sentencing outcomes across jurisdictions within a state and especially at the impact of disparate sentences on offenders.

As a first step toward determining the applicability of sentencing guidelines on a statewide scale, LEAA's National Institute of Law Enforcement and Criminal Justice has decided to support a field test of guidelines on a multijurisdictional basis in the criminal courts of a few selected states. This test will make use of the methodology previously developed during the feasibility study in individual jurisdictions to establish a common set of sentencing guidelines across several jurisdictions within each state. Specifically, the major objective of the sentencing guidelines multijurisdictional field test design is to evaluate the effectiveness of sentencing guidelines as a mechanism for enhancing sentencing consistency across different jurisdictions within a state. The second objective, prerequisite of course to the major objective, is to test the feasibility of developing and implementing sentencing guidelines across a number of jurisdictions within a state. A third objective of the sentencing guideline test design is to provide a body of knowledge for jurisdictions looking for a means to structure judicial decisionmaking. The following sections of this document set forth the parameters of a multijurisdictional sentencing guidelines field test that should be developed and implemented in each state in order to assess the field test experience in terms of its objectives.

---

[7] Wilkins, et al., pp. xvi-xviii.

[8] Their impact, however, is not yet known. An evaluation to examine the effects of guidelines on sentencing variation within each of the four test sites is currently being conducted under the auspices of the LEAA.

II. SENTENCING GUIDELINES DEVELOPMENT AND IMPLEMENTATION

A. Overview

This document describes the essential processes to be undertaken for guideline development and implementation in the sites selected for participation in the field test of sentencing guidelines. This field test design is expected to be implemented in several jurisdictions in at least two states. Each selected state and its participating jurisdictions constitute a "site" for the purpose of this test design. Depending on the court structure in a state, jurisdictions in each site may be counties, judicial districts, or a combination thereof. However, it is expected that participating jurisdictions within each site will be of a heterogeneous nature: that is, some combination of urban, suburban, and rural jurisdictions, with some indication of disparate sentencing practices, will be selected. The number of jurisdictions within each site is not prescriptive and will depend on factors such as cost, guideline construction approach, and site selection.

The development of sentencing guidelines for this field test will be accomplished by utilizing a combined data base representing all participating jurisdictions in a site; this data base provides the empirical basis for a common set of guidelines to be used by each participating jurisdiction. While other approaches appear feasible (such as a process in which each jurisdiction in a site independently develops its own guidelines and then negotiates a common set with the other jurisdictions in that site), this approach is recommended because it minimizes the time and costs involved in developing guidelines on a multijurisdictional basis. At the same time, this approach retains the empirical basis needed to examine sentencing decisions both within and among jurisdictions in each site.

Figure 2, below, summarizes diagrammatically the field test approach for developing sentencing guidelines on a multijurisdictional basis using a pooled data base.

The methodology for the development of sentencing guidelines for each test site follows the basic approach documented by the CJRC, Albany, New York.[9] This methodology involves the following sequential tasks:

- data collection;
- statistical analysis of sentencing decision information;
- guideline model development;

---

[9] See Footnote 5.

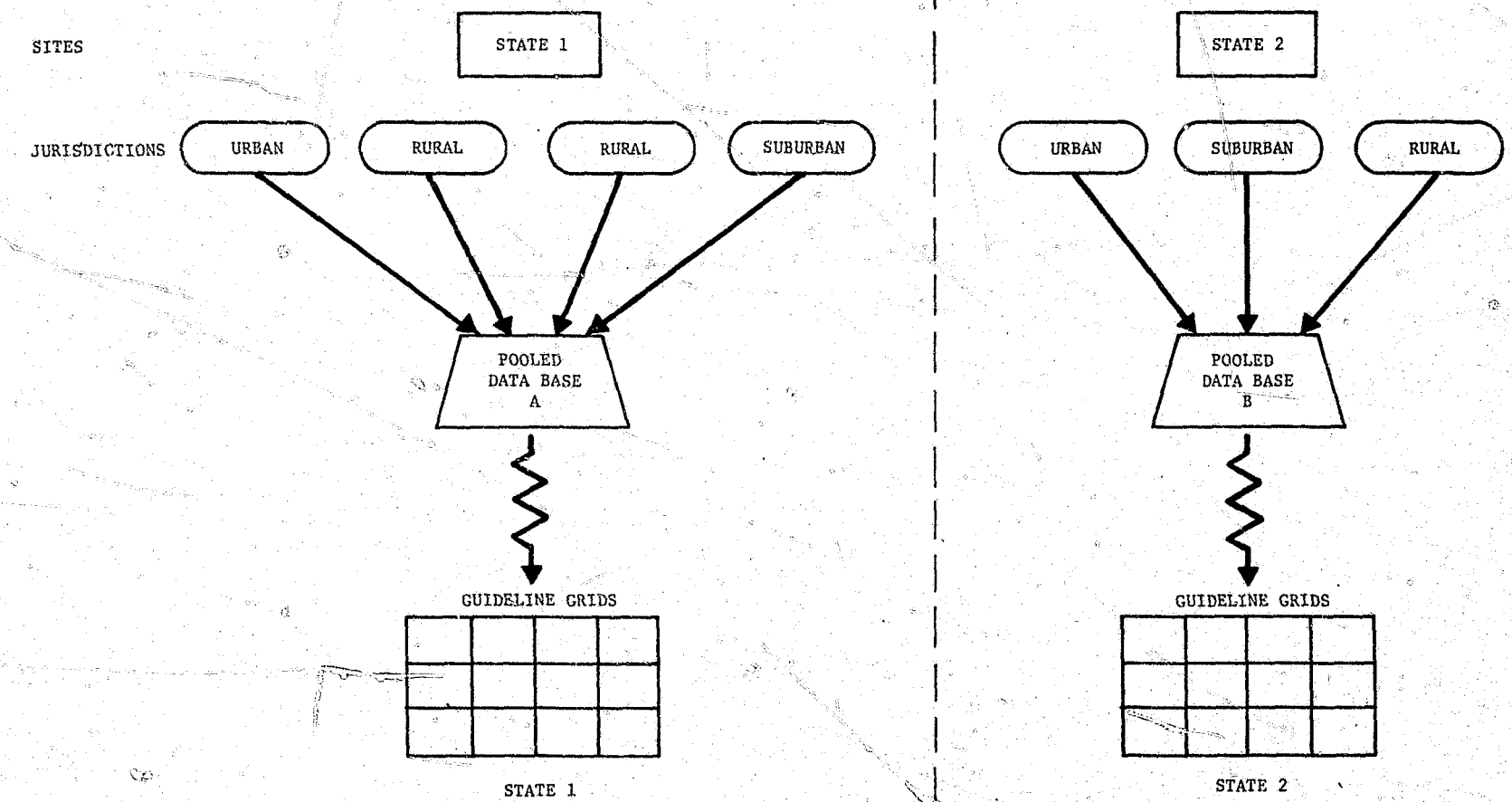

FIGURE 2
TEST DESIGN APPROACH FOR MULTIJURISDICTIONAL SENTENCING GUIDELINES

- guideline validation; and
- guideline use, review and modification.

The field test design is based on methods which have proven feasible in the development and implementation of sentencing guidelines to date. The approach represented by this field test design is in no way exhaustive and some aspects of it have been criticized. Therefore, the sentencing research team in each site may wish to explore alternative methods for developing sentencing guidelines. What is critical is that the various tasks be pursued in a sequential fashion for the purpose of producing an explicit tool for structuring judicial decisionmaking.

The process for developing sentencing guidelines outlined below (see Figure 3) requires close collaboration between those who develop the guidelines (project staff) and the judiciary, since guideline development is as much a result of judicial policy decisions as it is of empirical analysis. For example, the judiciary may insist that some factors be included in the guidelines even though they are not found to be as strongly associated with sentencing decisions as other factors. Similarly, judges may select a guideline model which may be less predictive but for policy reasons may be more desirable than alternative models. It cannot be overemphasized that successful guideline development and implementation rest upon the willingness of the judiciary to make policy decisions based on this collaborative effort.

The development of sentencing guidelines begins with activities related to collecting data that will be used to determine the implicit sentencing policy of each site. These activities include: an analysis of each site's criminal code and the sentencing practices of their courts; the design of a coding manual and an instrument to collect specific types of information considered most important for guideline development; and the selection of a construction sample to provide the data base needed for analyzing sentencing decisions and for constructing sentencing guideline grids.

Statistical analyses of the sample data comprise the second major set of activities related to the development of sentencing guidelines. These analyses permit identification of those sentencing factors most related to sentencing decisions. Furthermore, statistical analyses of this set of factors provide the basis for weighting those factors selected for inclusion in the offender and offense scales used to construct guideline grids.

TIME FRAME	ONE (7 MONTHS)	TWO (5 MONTHS)	THREE (12 MONTHS)
STAGE	INITIAL PLANNING AND DATA COLLECTION →	MODEL DEVELOPMENT, ASSESSMENT AND SELECTION →	GUIDELINE IMPLEMENTATION AND OPERATION
TASKS	• HIRING OF PROFESSIONAL AND CLERICAL STAFF • ANALYSIS OF PENAL CODE, COURT SYSTEM, AND DATA SOURCES AVAILABLE TO JUDGES • DEVELOPMENT OF DATA COLLECTION PLAN FOR CONSTRUCTION SAMPLE • INITIAL DEVELOPMENT AND PRETESTING OF DATA COLLECTION INSTRUMENT FOR THE CONSTRUCTION SAMPLE • ESTABLISHMENT OF THE ADVISORY BOARD • REVIEW OF DATA COLLECTION INSTRUMENT AND PLAN WITH ADVISORY BOARD • HIRING AND TRAINING OF CODERS • COLLECTION OF DATA • HIRING COMPUTER PROGRAMMER • TESTING OF INTER- AND INTRA-CODER RELIABILITY • PREPARATION OF DATA FOR ANALYSIS: CLEANING, STORAGE AND CHECKING	• CONSIDERATION OF MODELS • ANALYSIS OF DATA • REVIEW OF RESULTS WITH ADVISORY BOARD • DEVELOPMENT AND TESTING OF PRELIMINARY GUIDELINE MODELS BASED ON CONSTRUCTION SAMPLE • PRESENTATION OF RESULTS TO ADVISORY BOARD AND SELECTION, SYNTHESIS OR DEVELOPMENT OF ADDITIONAL MODELS (AS APPROPRIATE) FOR FURTHER TESTING • DEVELOPMENT OF DATA COLLECTION PLAN FOR VALIDATION SAMPLE AND DESIGN OF DATA COLLECTION INSTRUMENT • COLLECTION OF DATA • TESTING OF INTER- AND INTRA-CODER RELIABILITY • PREPARATION OF DATA FOR MODEL VALIDATION • VALIDATION OF ACCURACY OF PRELIMINARY GUIDELINE MODELS • PRESENTATION OF THE RESULTS TO ADVISORY BOARD, SELECTION OF ONE MODEL TO REPRESENT SENTENCING POLICY AND RESOLUTION OF ATTENDANT POLICY ISSUES • ESTABLISH, BASELINE LEVEL OF SENTENCING VARIATION IN TERMS OF ACCEPTED GUIDELINE MODEL o ESTABLISH THE CRITERIA FOR THE REDUCTION IN SENTENCING DISPARITY	• DEVELOPMENT OF THE DOCUMENTATION REQUIRED TO IMPLEMENT THE GUIDELINES, E.G., GUIDELINE MANUALS, GUIDELINE SCORESHEETS AND CODING MANUALS • ASSIGNMENT OF PERSONNEL TO SUPPORT THE GUIDELINE SYSTEM, E.G., TO PREPARE AND COLLECT THE GUIDELINE SCORESHEETS, ANALYZE THE DATA, CONDUCT REVIEW SESSIONS AND MODIFY THE GUIDELINES WHEN REQUIRED • ACQUISITION OF INFORMATION BY JUDGES AND SUPPORT PERSONNEL CONCERNING THE USE OF THE GUIDELINES • USE OF THE GUIDELINES: - PREPARATION OF THE GUIDELINE SCORESHEET - JUDICIAL CONSIDERATION OF THE SUGGESTED SENTENCE - PROVISION OF REASONS FOR IMPOSITION OF A SENTENCE OTHER THAN THE ONE SUGGESTED BY THE GUIDELINES - COLLECTION OF GUIDELINE SCORESHEETS • ANALYSIS OF SENTENCING DATA WITH ADVISORY BOARD • THREE REVIEWS OF THE GUIDELINES TO ASSESS CHANGES IN SENTENCING POLICY • CONTINUOUSLY PROVIDE THE EVALUATOR WITH THE DATA NEEDED TO ANALYZE THE IMPACT OF GUIDELINES ON BASELINE VARIATION IDENTIFIED IN STAGE TWO

FIGURE 3

TIMETABLE AND TASKS FOR IMPLEMENTATION

The next major task in the development of guideline models involves constructing alternative sets of grids using various offender and offense scales comprised of weighted combinations of those factors found to influence sentencing decisions. These grids are then examined for their ability to predict the incarceration/non-incarceration (hereafter, "In/Out") aspect of the sentencing decision. The data for this examination are the same as that used to initially construct the guideline grids.

Once the best models (sets of grids) for predicting the "In/Out" decision are determined, it is necessary to test their accuracy on a validation sample of actual sentencing decisions. Following this validity check, the construction and validation samples are combined to provide historical data in order to develop normative sentence lengths for the "In" cells of the guideline grids. These models and the results of testing are made available to the judiciary in order to select one model that will be used as guidelines for their sentencing decisions. The selection of a guideline model for implementation will allow the identification by project staff of the baseline levels of sentencing variation within the jurisdictions. These baseline levels will later be used to determine the impact of guidelines on variation. It will also serve as the basis for establishment by the judges of the criteria for reducing disparity. Once a specific guideline model has been accepted by the judiciary of each site, procedures must be instituted to ensure that these guidelines are formally implemented. Formal implementation means that the judiciary should promulgate by court rule or special directive a requirement that guidelines be used. This directive should ensure that all judges consult the guidelines in conjunction with their sentencing decisions. Specific written reasons must be provided in all cases where the actual sentencing decision differs from that suggested by the guidelines. Formal implementation also involves establishing an organizational mechanism to support the use of the guidelines by the judiciary. This includes assigning responsibility for computing sentencing decisions according to the guidelines, collecting and analyzing information concerning the use of the guidelines, and presenting this information to the judiciary for review and possible modification of the guideline grids.

After an initial period of use, the guidelines are reviewed by the judiciary. This review focuses on a comparison of actual sentencing decisions with those indicated by the guidelines. Modification of the guidelines may be made to more closely reflect changes in the current sentencing practices or to accommodate explicit policy recommendations of the judiciary.

The subsequent sections of this field test design are organized to describe the requirements for establishing sentencing guidelines in a test site. These sections are as follows:

- project organization;
- data collection;
- model development; and
- model implementation.

As previously stated, Figure 3 provides the timetable and tasks for this test design. It should be noted that these tasks are sequential, that is, for the most part one task must be completed before the next one is begun.

B. Project Organization

There are a number of organizational requirements which appear to be important to the development and use of sentencing guidelines. Key among these are the organizational locus of the project, the project staff, technical help, and a judicial advisory board in each participating site.

The project staff for developing sentencing guidelines in each site should consist of at least two full-time members: a project director and a research director. The project director will be responsible for the day-to-day administration of project activities, and also serve as the project's primary contact with the advisory group of the judiciary designated to make policy decisions about guideline development and use. Among other skills, the project director should possess a working knowledge of the site's judicial system, its sentencing practices and, if possible, its judicial personnel. A member of the existing staff of a state-level court organization (e.g., State Court Administrator's Office) possessing these skills would thus be a likely selection. The research director will be responsible for overseeing the technical development of sentencing guidelines. This person will either personally conduct or supervise others (data collectors, computer programmers, keypunchers) in the data collection, statistical analyses and model building activities required for guideline development.

While the project director and the research director will comprise the core staff in each site, it will be necessary to supplement this staff throughout the project. Data collectors, coders, computer programmers, keypunchers, and clerical staff will be needed to complete each project's activities.

It is also likely that guideline development in each site will require outside assistance by specialists in particular technical areas. For example, it will be necessary to obtain help to brief the project staff and the judiciary concerning the conceptual basis of guidelines prior to undertaking specific project activities. Determining the size of the sample of sentencing decisions on which to base guideline development, conducting analyses of sentencing decisions, and developing alternative guideline models are other project tasks which are likely to require outside assistance.

An Advisory Board representing the judges from the participating jurisdictions will be established for each test site. The purpose of this Board is to oversee the project staff's activities and to make those policy decisions needed to develop sentencing guidelines. The Advisory Board should consist of no more than 10 judges, all of whom should be experienced and currently involved in setting sentences in criminal cases. The composition of the Board should reflect the relative volume of sentencing decisions among the participating jurisdictions in each site. For example, if 4,000 sentencing decisions were made in an urban jurisdiction, 2,000 made in a suburban jurisdiction(s), and 1,000 in a rural jurisdiction(s), the composition of the Board might be as follows: four urban judges, two suburban judges, and one rural judge. Because of the nature of the activities of the Advisory Board, it is important that its members be delegated the appropriate authority and responsibility to represent their constituency regarding policy decisions.

It may be desirable that other individuals be included as ex-officio members of each site's Advisory Board. To keep the Advisory Board at a manageable size, the number of ex-officio representatives should be limited to approximately six additional members. These individuals may include: judges from non-participating jurisdictions; representatives of prosecutorial, public defender, court administrative, and correctional agencies within the participating jurisdictions; citizen representatives; and a consulting expert in research methodology. The latter member would provide the Board with a review capability concerning the project's technical activities.

In summary, it should be emphasized that the nature of the interactions between the project staff, the Advisory Board, and the body of judges in the participating jurisdictions will largely determine the extent to which sentencing guidelines will be successfully developed and implemented in each site.

C. Data Collection

After the initial meeting between the Advisory Board and the project staff, the first major step in developing sentencing guidelines involves establishing a data base from which the guidelines will be constructed. Data collection and guideline development are based on the belief that an accumulation of case-by-case sentencing decisions incrementally results in an implicit sentencing policy. Further, it has been demonstrated that this implicit policy can be made explicit through an empirical analysis of those factors commonly held to be most influential in sentencing decisions.[10] Sentencing data provide the basis upon which the implicit sentencing policy of the courts in each project site will be described.

There are a number of factors that must be considered that will shape the data base for developing sentencing guidelines. Broadly defined, these are the statutory framework and criminal code which define the substance and procedures related to sentencing practices; the nature of the guideline models to be developed; and the type of information upon which judges make sentencing decisions in each site.

A complete review of each site's criminal code must be completed prior to determining the type of guidelines that can be developed and the kinds of sentencing information to be collected. The guidelines must be constructed within the legislative mandates that define sentencing practices and the range of dispositional alternatives available to judges. For example, the relative seriousness of various offenses, minimum and maximum penalities prescribed by law, and the exclusion of specific factors by law from consideration in sentencing will probably influence data coding decision rules and the eventual form of the guidelines.

The types of guideline models to be developed by each site will be a critical consideration. Particular types of models (see Section II D 1) and the nature of the sentencing decision reflected in a model will affect the selection of information items, the size of the data base and the data analysis. To illustrate, sentencing decisions can be viewed as a bifurcated process: 1) whether or not to incarcerate an offender; and 2) if incarcerated, the time or length of the particular type of sanction. It is conceivable that the factors which affect "In/Out" decisions are different from those which influence decisions concerning length of incarceration. This provides two options for selecting the kinds of guideline models to be developed.

---

[10] Wilkins, et al., p. 10.

The first option is one that combines the bifurcated decision in a single grid. This grid is developed in terms of an analysis of factors which influence the "In/Out" decision. After the grid is constructed, sentencing ranges based on prior experience are calculated for the cells of the grid. The second option involves developing separate grids for each aspect of the bifurcated decision. This necessarily involves analyzing separately the influence of factors on both the "In/Out" decision and sentence length.

The choice between these options will have significant ramifications in terms of sample size and the number and type of sentencing factors upon which data are gathered. For example, in order to conduct a statistically valid analysis of those factors most useful in describing sentence length, the sampling would have to be conducted so as to ensure an adequate representation of "In" decisions. Although there are these two possible approaches for guideline development, this field test design encourages the use of a single grid approach described above, as documented by the CJRC site experiences.[11] The reasons for this recommendation are fourfold: 1) the relative simplicity of the single grid approach; 2) its practicability in terms of use by the judiciary; 3) its lower developmental costs; and 4) its proven feasibility in other sites.

Data collection activities also must be preceded by a comprehensive examination of the types of information available to judges to make sentencing decisions. This examination defines those factors which are candidates for inclusion in the empirical analysis of the relationship between various information items and the sentencing decision. The process of identifying candidate variables can vary among the test sites and the jurisdictions within sites. In one jurisdiction or site, all the sentencing information available to judges may be included in presentence investigation reports. In other sites, some of this information may be dispersed among written reports, verbal presentations to the judges in presentence or prepleading conferences, and other sources. In any case, the availability of standardized presentence investigation reports facilitates an efficient data collection effort. It should be noted that only information items common to all jurisdictions within a site should be included in the empirical analysis.

These sources of information can provide hundreds of variables with some possible influence on sentencing decisions. Additionally, an item of information may be defined differently not only in the

---

[11] Wilkins, et al., pp. 7-32; see also Gelman, et al., p. 15.

various sources of data, but also in the jurisdictions that comprise each site. The larger the number of variables to be examined, the more complex and costly the analysis is likely to be:

> Each item of information collected will add certain costs to the project in terms of time needed to collect, keypunch and analyze the data. Therefore, one must be cognizant of such initial pragmatic concerns in deciding just what information is needed.[12]

Data analysis conducted in previous sites suggest that of the large number of variables that might influence sentencing, only a small number have a statistically significant relationship to the sentencing decision.[13] Additionally, the variables found to be most predictive of sentencing decisions were similar regardless of the site. Because of these findings this test design recommends that data collection be restricted to approximately 30-50 variables reflecting:

- previous site experience;
- specific variables required by each test site's Advisory Board; and
- needs of the evaluation.

As noted, there is a core of common factors which have been found either to be predictive of sentencing decisions or of particular interest to the judiciary in those sites which have developed sentencing guidelines. Table I lists 48 of these variables as a starting point for each test site's choice of data items. In addition to this list, previous site documentation can be an extremely useful source of information for identifying factors for data collection.[14]

---

[12] Gelman, et al., p. 3.

[13] Wilkins, et al., pp. 10-15.

[14] Wilkins, et al., pp. 10-15, 44-50; and, Gelman, et al., pp. 27-70. Analysis of the data from previous sites can provide more complete information about the predictive ability of a wide range of factors. Unfortunately, these data are not currently available.

TABLE I

SUGGESTED LIST OF CORE VARIABLES

DISPOSITION DATA	• Type of sentence • Length of probation • Length of incarceration - minimum • Length of incarceration - maximum • Basis of adjudication
INSTANT OFFENSE DATA	• Number of offenses charged • Most serious offense charged • Number of offenses at conviction • Most serious offense at conviction • Number of offenders • Weapon usage • Type of weapon involved • Number of business victims • Number of personal victims • Physical injury suffered by victim • Value of property involved in offense • Involvement of drug distribution • Value of drugs • Description of drug involved • Offender's use of alcohol or drugs at time of instant offense
CRIMINAL HISTORY DATA	• Criminal status of offender at time of instant offense • Number of prior juvenile arrests • Number of prior juvenile convictions • Number of prior juvenile probations • Number of prior juvenile probation revocations • Number of prior juvenile incarcerations • Number of prior juvenile paroles • Number of prior juvenile parole revocations • Number of prior adult arrests • Number of prior adult convictions • Number of prior adult probations • Number of prior adult probation revocations • Number of prior adult incarcerations • Number of prior adult paroles • Number of prior adult parole revocations
DEMOGRAPHIC AND SOCIAL HISTORY DATA	• Offender's date of birth • Offender's sex • Offender's race • Offender's marital status • Offender's number of dependents • Offender's support of dependents • Offender's residential stability • Offender's work status • Length of employment • School status • Level of education • Offender's use of alcohol • Offender's use of drugs

It is likely that the members of the Advisory Board will have preferences about having certain items of information analyzed in terms of their impact on sentencing. Members of the Board may desire the inclusion of factors in the analysis because of their perception that a particular factor affects their own decisions. Others may see the project's data collection tasks as a unique opportunity to collect information for descriptive purposes regardless of this information's potential use in the guidelines. Both concerns are legitimate for shaping the types of information to be collected.

Evaluation needs may also dictate that certain items of information be included in the project's data collection effort. For example, the evaluator may want project staff to collect court processing data in conjunction with that collected for guideline development. Such requests will also play a part in shaping the project data collection effort. Additionally, for this test design, the data must be coded to allow identification of cases by judge and jurisdiction. Both factors may constitute sources of variation in sentencing practices and it is reasonable to expect that guidelines would reduce these disparities. Therefore, the variables "judge" and "jurisdiction" should be examined as part of the evaluation.

After decisions are made about the items of information to be collected, a coding manual will be developed for each test site. This manual will serve as a basic instructional tool for the project by identifying in clearly delineated, operational terms the variables to be collected. Decision rules concerning the format for coding data and the categorizations of ambiguous or unclear information should be made explicit in this manual. Two types of decision rules will be important in the development of the coding manual. The first type will involve decisions of a general nature that apply to the majority of variables comprising the collection effort. For example, it is necessary to establish specific values to distinguish between cases in which a characteristic or attribute is not present ("missing value"), from those cases in which it could not be present ("not applicable"), and from those cases in which a category of a variable is present but not listed in the coding instrument ("other"). Table II presents an example of general instructions to coders which result from such decision rules using the variable "Number of Prior Adult Probation Revocations."

TABLE II

CODING INSTRUCTIONS FOR AN EXAMPLE VARIABLE:
NUMBER OF PRIOR ADULT PROBATION REVOCATIONS

0 = Previously placed on probation, but no prior revocations
1-5 = Number of revocations
6 = Six or more revocations
7 = Not applicable, never placed on probation
8 = Other
9 = Missing value

A second type of decision rule to be reflected in the coding manual involves resolving policy issues implied by the use of individual variables. These decisions may have significant impact on the empirical basis of the guidelines and could be a source of controversy in their eventual acceptance by the judiciary. Because of this the responsibility for many of these decisions should rest with the Advisory Board of each site. For example, the collection of information about an offender's prior criminal history is one of many areas where a decision rule with obvious policy implications can occur. The issue of "decay" is an example of an important consideration in designing coding instructions. Decisions of whether to exclude the coding of prior offenses after the elapse of a certain time period (e.g., 20 years); or to code only particular kinds of crimes after some period of time (e.g., murder); or to count all criminal activities without placing limits on the time interval can have a measurable bearing on the outcome of the analysis for guideline development. There are other policy issues that may arise when putting together the coding manual. Documentation of previous guideline development experiences provides a comprehensive discussion of these issues and describes how they were addressed.[15]

A coding sheet should be developed once all the variables have been defined in the coding manual. The coding sheet should be constructed to reflect the order of information contained in the case file and/or other sources of data. The design of the coding sheet should promote efficient data collection and reduce the potential for coding error.[16] In all instances, the coding sheet should be pretested in each participating jurisdiction within a site on a minimum of 25 cases drawn randomly from the data sources. If necessary, the coding sheets should be modified to enhance data gathering activities.

---

[15] Gelman, et al., pp. 6-12; see also Calpin, et al.

[16] For an example of a well designed coding sheet, see Gelman, et al., pp. 71-81.

Drawing a sample of cases (the construction sample) to both examine the relationship of selected variables to sentencing decisions and to construct preliminary guideline grids constitutes the next phase of guideline development for the test design. The size of the sample will depend on a number of interrelated factors, including:

- the total number of sentencing decisions made in the participating jurisdictions during a specified sampling frame;

- the number of cases needed to satisfy the requirements of statistical techniques (e.g., correlation and regression) used to examine the relation between various offender and offense variables and sentencing decisions;

- the size of the sample needed to minimize the number of empty cells (or cells with very few cases) in each grid of a guideline model; and

- the time and cost required to gather data.

The minimum number of cases the judiciary is willing to use as a basis for establishing guidelines and the incarceration rate among the jurisdictions in each site are important considerations in estimating sample size. The size of the sample for each site must be large enough to provide sufficient experiential data within the cells of the guideline grids. The determination of what represents "sufficient" data will reflect the statistical concern for the reliability of the information and the judgment of each site's Advisory Board. The incarceration rate for each jurisdiction must also be examined to ensure that a sufficient number of "In" decisions are sampled to allow for a sound description of the time dimension of the sentencing decision.

In each of the previous guideline sites, a model, comprised of 3 to 6 grids consisting of 20 to 50 cells each, was developed. The developers of guidelines in these sites have recommended that a sample of at least 1,000 cases is required for statistical analysis, and that a larger sample (e.g., from 2,000 to 4,000) is needed to minimize empty cells in the grids.[17] However, it is important to note that this sample was based on a guideline development process in a single jurisdiction. The multijurisdictional development process in this test design may require a larger sample. Unfortunately, there is no

[17] Gelman, et al., p. 11.

definitive answer concerning sample size that can be prescribed prior to data collection in this setting. Therefore, technical advice should be sought to address the issue of sample size and to provide various options available to meet each site's needs.

In drawing the construction sample, a decision must be made concerning the type of sampling technique to use. This test design recommends that the size of samples drawn from the population of sentencing decisions in the participating jurisdictions in each site be representative of the relative caseloads of those jurisdictions. While other approaches (e.g., equal size sampling of cases from each jurisdiction) are also feasible, the representative sample provides the most historically accurate description of the collective sentencing experience of each site. Equal size sampling, on the other hand, could have some practical advantages in evoking maximum cooperation from the judges in the participating jurisdictions--especially when one jurisdiction dominates the others in terms of case volume. This problem of over-representation of one type of jurisdiction (e.g., the urban jurisdiction) can be reduced when selecting the combination of jurisdictions constituting each site. For example, it may be advisable to consider multiple rural or suburban jurisdictions. This would assist in balancing out the large volume of sampled cases in the urban site. To illustrate, one state may be represented by one urban, two suburban and no rural jurisdictions, while another state may be represented by one urban, one suburban and two rural jurisdictions (see Figure 2).

The time frame of the construction sample should be as current as possible. The more current the time frame, the more likely it is that the sample will reflect current court policy. Further, the sampling frame should be selected to minimize any abnormal variations that may occur as a consequence of seasonal effects. Therefore, this test design recommends that the sample frame cover the most recent 12-month period of available case data.

The hiring and training of coders is central to a successful data collection effort. Data coders should, if possible, be selected from persons familiar with criminal justice records and legal terminology. Law students and students in the social sciences would be appropriate candidates. The number of coders required at each site will depend upon the physical location of the data sources, the quality of these sources, the size of the sample, the number of variables to be collected, and the amount of time coders are available. Assuming a construction sample of approximately 4,000 cases, a decentralized data base and case records of average quality, it would require 15 to 20 coders, each working half-time (20 hours per week), approximately 15 weeks to collect the construction sample in each site.

Data coders must be well trained by project staff in the use of both the coding manual and coding sheets. To familiarize the coders with the terminology and format of the data sources, each coder should code a small common set of cases. In this way, general problems can be identified and resolved prior to collecting the construction sample.

After the data collection instrument has been designed and pretested, the sample size, time frame and sampling technique decided, data collection can begin. Project staff must closely supervise the data collection effort. This supervision includes monitoring coders to detect individual or common problems, allocating cases for coding in an efficient manner, and conducting inter- and intra-coder reliability tests at regular intervals. Additionally, as coders turn cases in, project staff must review each coding sheet to ensure that all items have been completed and that there are no obvious inconsistencies in coding.

Concurrent with data collection, information should be keypunched and verified. After all the data have been compiled on cards, the data should be "cleaned," that is, searched for mistakes and corrected prior to analysis. There are several methods for cleaning the data: 1) visual inspection of raw data as printed from the cards; 2) examination of frequency distributions for inconsistencies; and 3) use of error statements to check the internal consistency of the data.[18]

D. Model Development

The actual development of a sentencing guideline model for use by the judiciary "is an iterative process of testing, modification, and retesting."[19] The construction sample data previously described provides the information base needed to build the grids of guideline models. These grids will be two-dimensional, consisting of an offender and offense scale.

This developmental process involves a series of sequential activities. These activities include:

- consideration of model types;
- statistical analysis of the construction sample;

---

[18] Sample error statements and methods used to clean data are explained in detail in Gelman, et al., pp. 12-13, and Appendix C.

[19] Gelman, et al., p. 18.

- assignment of weights to the selected sentencing variables;

- development and evaluation of alternative guideline models using the construction sample;

- modification and selection of models for validation by the Advisory Board; evaluation of selected models with a validation sample; and

- selection and adjustment of a guideline model for use by the judiciary.

This series of activities can be seen as a filtering process--that is, from many possible variables and sets of grid configurations based on these variables--one model using a small set of variables is agreed upon by the judiciary as a representation of their sentencing policy (see Figure 4).

1. Construction of Model Types

The choice of particular types of guideline models (for potential selection by the Advisory Board) reflecting various possible categorizations of offenses will have important ramifications for the analysis as discussed earlier. There are at least four types of guideline models which can be employed. These are:

- unitary models that develop one grid for all of the specific types of criminal offenses;

- statutory models that develop specific grids to conform with various statutory classifications of crime; this could be as simple as a misdemeanor/felony dichotomy or as detailed as the statutory classifications of a criminal code (e.g., Felony One, Felony Two, etc.);

- generic models that develop specific grids to conform with various offense types (e.g., property, violent, and drugs); and

- crime-specific models that develop grids for each crime (e.g., burglary, robbery, etc.).

At this point in guideline development, the selection of some types of models as feasible and/or desirable and the exclusion of others is critical because data analysis will be conducted in terms of potential model types. For example, the unitary model would allow

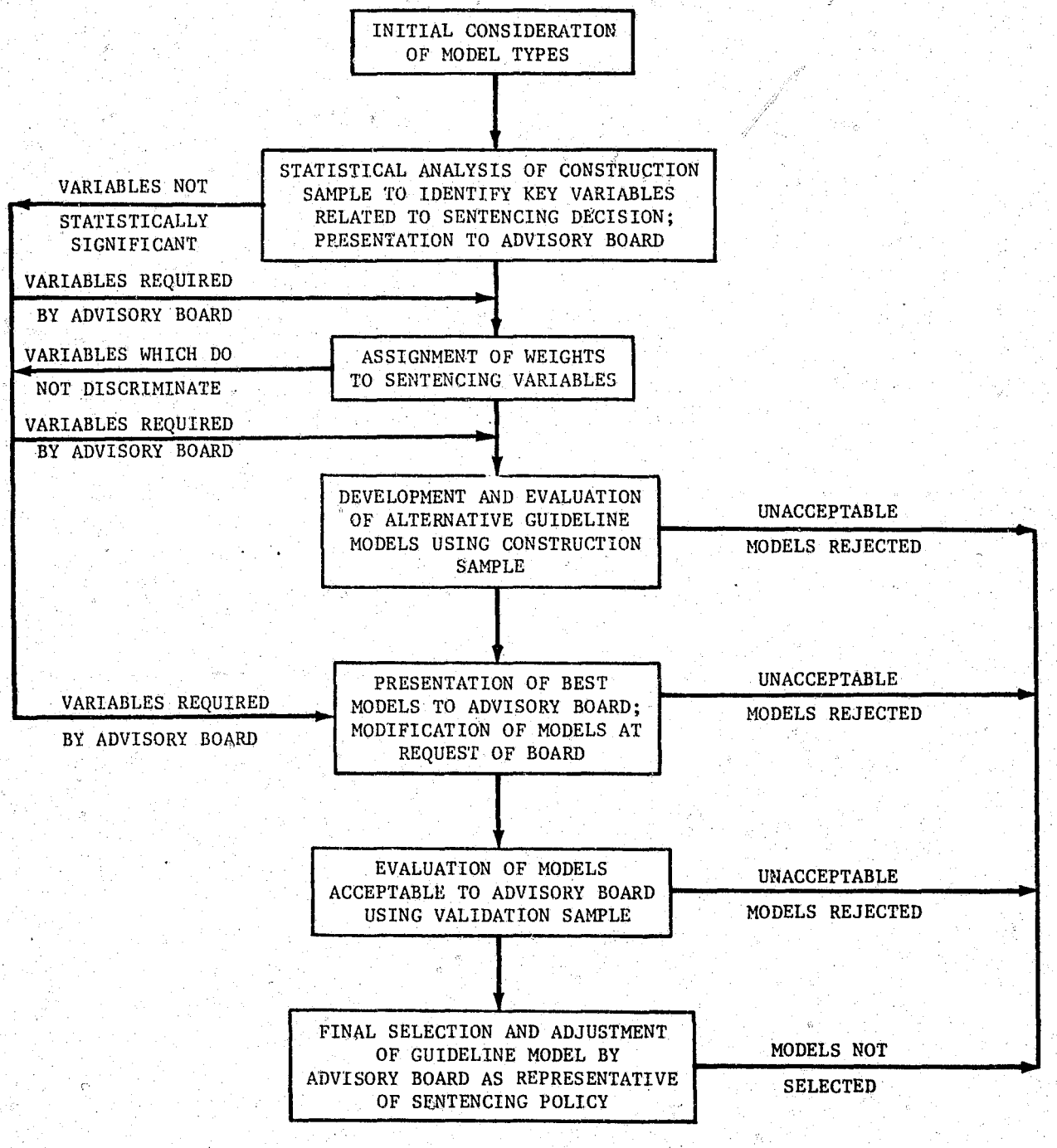

FIGURE 4
SENTENCING GUIDELINES
MODEL PROCESS DEVELOPMENT

the data analysis to be conducted over the total sample. The major advantage of the unitary model is that it requires the simplest analyses because only one set of independent variables (i.e., information items) is considered for the entire sample of cases representing all offense types. Each of the three other model types--statutory, generic, or crime-specific--necessarily requires more data analysis since a number of separate grids are constructed for any of these models. Because these models make a finer distinction among crime types, they allow a more accurate mapping of sentencing decisions and of the particular variables influencing these decisions. For example, selecting the generic model requires that subsets of the construction sample, based on the generic categories of crime (e.g., property, violent, etc.), be analyzed separately so that the information items influencing sentencing decisions for each generic category can be identified.

This test design recommends that either a statutory or generic model be developed in each site. Although the unitary model requires fewer analyses, it is likely that this model will not provide a sufficiently complete and accurate representation of sentencing policy to satisfy the judiciary. Conversely, the crime-specific model, while capable of providing a more complete and accurate representation of policy, requires such extensive data collection and analysis as to be prohibitively expensive.

The selection of potential model types will involve a consideration of each site's criminal code and the particular preferences of the Advisory Board. For example, in some states the statutory classification of offenses may be so unstructured (e.g., each offense has a unique sanction prescribed by the legislature) that the development of a statutory model may be difficult. In other states, the prescribed sanctions for certain statutory classes (e.g., Felony Three and Felony Four) may be so similar as to allow their combination in one grid.

Prior to data analysis, it is only necessary to determine which model types may be feasible and acceptable to the judiciary. The final selection of a particular model will be based on model testing and validation. It is anticipated that this model will be either of the statutory or generic type. In either case, this implies the development of two to six sentencing grids reflecting offense categories. The cells within these grids will contain, as a minimum, both a recommended "In/Out" decision and a sentence range for those incarcerated.

2. Statistical Analysis

After a consideration of model types, model development continues with the statistical analysis of the construction sample data in order to identify those variables most predictive of sentencing decisions.

It should be noted that the various analyses described in this section are performed on subsamples of the construction sample. These subsamples are to be based on the potential model types. For example, the development of a generic model requires statistical analyses to be conducted for separate subsamples of violent offenses, property offenses, and drug offenses.

Data analysis should begin with the description of the frequency distribution of each of the variables for the total sample and appropriate subsamples. These distributions will later assist the project staff in making decisions concerning the recoding of the data and the choice of the statistical analyses upon which model development is based.

Crosstabular analysis is the first of three statistical techniques recommended for examining the relationship between various sentencing variables and the sentencing decision. In this analysis, the sentencing variables agreed upon by the Advisory Board (Section II C) constitute the independent variables. The sentencing disposition, defined as either "In" or "Out," is the dependent variable. In performing the crosstabular analysis, decisions must be made regarding the recoding of categories of variables. Outlying values and such categories as "not applicable" or "other" should be collapsed to facilitate the analysis.

Recoding will, of course, affect the analysis and may, like the original decisions about coding (Section II C), also influence policy. Based on previous experience, decisions about the recoding of variables dealing with criminal history information and the nature of criminal offenses should be made with the advice of each site's Advisory Board. Tables III and IV provide an example of a crosstabulation where qualitatively different categories are collapsed to facilitate data analysis.

The dependent variable for the crosstabular analysis may also present problems in terms of recoding categories. A simple "In/Out" dichotomy may be complicated by alternative sentence dispositions such as work release, community corrections, or "split" sentences. As with the independent variables, recoding categories of the dependent variable may affect analysis and policy. For example, previous site experiences suggest that:

TABLE III

CROSSTABULATION OF OFFENDER'S
RELATIONSHIP TO CJS (BEFORE COLLAPSING)

NATURE OF RELATIONSHIP

SENTENCE TYPE	FREE	FREE CRIMINAL ACTION PENDING	JUVENILE PROBATION	ADULT PROBATION	JUVENILE PAROLE	ADULT PAROLE	INCARCER-ATED	ESCAPEE	OTHER *	MISSING VALUE	ROW TOTAL
OUT	430 (80%)	24 (60%)	15 (25%)	25 (25%)	6 (20%)	0 (0%)	0 (0%)	0 (0%)	20 (66.7%)	30 (75%)	600 (60%)
IN	120 (20%)	16 (40%)	45 (75%)	75 (75%)	24 (80%)	60 (100%)	30 (100%)	10 (100%)	10 (33.3%)	10 (25%)	400 (40%)
TOTAL	600	40	60	100	30	60	30	10	30	40	1000 (100%)

*OTHER represents individuals mandatorily hospitalized or under some other type of legal control outside the CJS.

TABLE IV

CROSSTABULATION OF OFFENDER'S
RELATIONSHIP TO CJS (AFTER COLLAPSING)

NATURE OF RELATIONSHIP*

SENTENCE TYPE	NO CJS SUPERVISION	CJS SUPERVISION	TOTAL
OUT	554 (78%)	46 (15.8%)	600 (60%)
IN	156 (22%)	244 (84.2%)	400 (40%)
TOTAL	710	290	1000 (100%)

*In this example, four categories--"free," "criminal action pending," "other," and "missing value"--have been collapsed into one category--"no CJS supervision." The six other categories have been collapsed into one category--"CJS supervision."

> judges may indicate that a certain sentence is considered an 'in' while the data indicates that it more closely resembles an 'out' decision. Consequently the judges will either have to make an early policy decision on how to classify certain sentences, or the analysis and model development will have to be done in duplicate with the dependent variable dichotomized in two or more different ways.[20]

Once the data have been recoded, correlation coefficients can be computed as a supplemental method to the crosstabular analyses. These computations are used to further examine the direction, strength, and significance of the relationship between each independent variable and the dependent variable (the "In/Out" decision). Measures of association such as Pearson's r, Kendall's tau, and Spearman's rho can be used. Variables that are highly intercorrelated or exhibit a low correlation with the "In/Out" decision may then be excluded from further analysis and model testing.[21]

Multiple regression analysis is the third statistical technique suggested for examining the relationship between sentencing variables and the "In/Out" decision. Multiple regression analysis of those variables not dropped on the basis of the crosstabular or correlational analyses serves two purposes.

First, multiple regression can address the problem of multicollinearity (i.e., two highly intercorrelated variables that account for the same variance in a dependent variable). For example, both prior arrests and prior convictions may be significant predictors of the "In/Out" decision. However, they are likely to be highly intercorrelated. Multiple regression allows the examination of the independent contribution of one variable to the prediction of the "In/Out" decision after the other variable has been considered.

Second, multiple regression provides an indication of the increase in predictive power gained by adding individual variables to the prediction of the "In/Out" decision. For example, in previous sites it was found that no significant increase in the prediction of the "In/Out" decision was gained after approximately the first ten independent variables had been used.[22]

---

[20] See Gelman, et al., p. 14.

[21] Ibid., for recommendations concerning the handling of missing information in these analyses.

[22] Calpin, et al., pp. 10-14.

This test design has recommended the development of a single grid approach (see Section II D 1) based on an examination of factors influencing the "In/Out" decision and not sentence length. Thus, the analyses described above employed "In/Out" as the sole dependent variable. In this approach, sentence length is determined by plotting actual prison terms in the appropriate cells, analyzing the results, and establishing sentence ranges based on the historical data. If it were decided that grids were to be developed to be predictive of sentence length, these same analyses (described above) would need to be conducted using "length of incarceration" as the dependent variable.

With the completion of the statistical analyses, project staff should have a good idea of those variables most closely related to the "In/Out" decision. A presentation should then be made to the Advisory Board regarding these findings. At this presentation, project staff should inform the Board which variables were and were not found to be related to the sentencing decision, the strength and direction of these relationships, and the interrelationships among independent variables. This presentation allows the Board the opportunity to decide which variables will be further tested for inclusion in the guidelines. It should be noted that the Advisory Board may desire that certain variables, although not found to be statistically significant in terms of their relation to the sentencing decision, be incorporated for further model testing.

3. Assignment of Weights to Sentencing Variables

The next step in guideline development involves the process of assigning weights to those variables selected for further testing so that offense and offender scales can be constructed. Each sentencing grid will be two-dimensional--that is, it will have two scales: one comprised of different combinations of the variables focusing on the offense, and the other comprised of combinations of variables dealing with the offender. The categorization of scores for each scale will determine the number of cells within any grid.

Before the project staff can proceed with the weighting of variables, the Advisory Board must make a policy decision about the issue of offense seriousness. The judicial members of the Advisory Board are likely to want a more precise representation of offense seriousness than the statutory definition of the offense at conviction provides. Previous guideline experience and research have found that judges take into consideration the "real offense" rather than merely the offense at conviction in deciding the sentence to impose.[23] "Real offense" is defined as the judicial perception of the actual behavior

[23] Wilkins, et al., pp. 26-27.

that characterized the commission of the crime by the offender. The "real offense" naturally appears to be most important to the judiciary in cases in which the offense at conviction represents a plea bargain.

For this reason, a key issue to be decided by each site's Advisory Board is whether or not to incorporate information reflecting the "real offense" into the offense scale of the guidelines. Regardless of the decision, a hierarchy of offense seriousness must be established for the classes of crime reflected in the guideline models. One method for establishing an offense hierarchy is to examine the construction sample data to determine the actual sentences imposed for each offense.* This information will provide a rank ordering of offenses reflecting implicit judicial policy. Another method for developing a hierarchy is a modified version of the Q-sort.[24] This method, which involves having the judiciary establish offense rankings on the basis of their own perception of seriousness, is summarized as follows:

> ...take offenses within a given category and classify them by the perceived seriousness of the typical offense. The rankings are established by the judges themselves...it is a fairly simple procedure in which the judges lay out cards in front of them and then group the cards in comparison to one another. After the rankings have been tabulated, the results are reported to the judges who, as a group, should resolve any disagreements or make any modifications...[25]

It should be noted that the Q-sort approach can be used in conjunction with the examination of actual sentences in the construction sample data to devise a ranking of offense seriousness.

The Advisory Board may decide to use the techniques above to establish an offense scale based only on the offense at conviction. On the other hand, the Advisory Board may decide that the guidelines should reflect the "real offense." In this case, the offense scale may be based solely on a ranking of "real offense" behavior or on a ranking

*Actual sentence severity will include elements for prior record as well as the seriousness of the instant offense. Such interaction should be examined.

[24] For further instruction regarding this method see Gelman, et al., pp. 17-18; also see, Gottfredson, et al., pp. 70-76.

[25] Gelman, et al., pp. 17-18.

of offenses at conviction as modified by factors relating to the real offense. Seriousness modifiers may reflect such information as weapon usage, degree of physical harm, or sale of drugs.

Once the policy decision about offense seriousness has been made, project staff can proceed with the development of the offense and offender scales. To accomplish this task, specific weights must be assigned to each variable considered for inclusion in the scales. This test design recommends a Burgess type system for assigning weights to the variables of the offense and offender scales. While there are other methods that could be used (e.g., assigning beta-weight coefficients from a regression equation), this method is recommended because of its predictive ability with criminal justice data and its computational simplicity.[26] The Burgess system assigns unit weights (points) to categories of independent variables associated with a dependent variable. The weighted variables are then combined to form a predictive scale.

The assignment of Burgess type weights to the variables being considered for the offense and offender scales is based on a reexamination of the earlier crosstabular analysis. The goal of this task is to recode the independent variables so that the rate of incarceration for each category of an independent variable differs from every other category of that variable and from the base incarceration rate. Categories should be developed which maximize differences in rates of incarceration. Table V presents an example of the Burgess weighting of one variable, "prior adult felony convictions." Assume that the base incarceration rate of the construction sample is 40 percent in a particular site. An examination of the crosstabulation between the "In/Out" decision and the number of prior adult convictions reveals that 15 percent of the offenders with no prior convictions are incarcerated. For those with one conviction, the rate of incarceration is 55 percent; for those with two or more, 70 percent. Thus, a logical recategorization of this variable for the assignment of Burgess type weights might be as follows (see Table V):

- no prior adult felony conviction;
- one prior adult felony conviction; and
- two or more prior adult felony convictions.

6

[26] See Gottfredson, et al., pp. 44-49 for an assessment of the Burgess method and additional references; see also Calpin, et al., pp. 20-22.

TABLE V

BURGESS WEIGHTING OF
PRIOR ADULT FELONY CONVICTIONS

NUMBER OF PRIOR CONVICTIONS

SENTENCING DECISION	0	1	2	3	4	5	6	7	8	9	10 or MORE	TOTAL
OUT	170 85%	45 45%	10 40%	4 20%	6 30%	3 25%	8 64%	2 20%	0 0%	0 0%	0 0%	248 60%
IN	30 15%	55 55%	15 60%	16 80%	14 70%	9 75%	5 36%	8 80%	2 100%	3 100%	5 100%	162 40%

(ORIGINAL CROSSTABULATION)

NUMBER OF PRIOR ADULT FELONY CONVICTIONS

SENTENCING DECISION	0	1	2 or MORE	TOTAL
OUT	170 85%	45 45%	33 30%	248 60%
IN	30 15%	55 55%	77 70%	162 40%

(AFTER RECATEGORIZATION)

NUMBER OF PRIOR ADULT FELONY CONVICTIONS	0	1	2 OR MORE
BURGESS WEIGHT	0	+1	+2

(BURGESS WEIGHTING)

The predictive efficiency of the recategorization of this variable as well as any recategorization of the other variables can be tested by a variety of methods including the Index of Predictive Efficiency and Mean Cost Rating.[27] These methods provide measures of the discriminatory power (in terms of the "In/Out" decision) of any particular categorization.

Once the variables have been recategorized to achieve maximum discrimination in terms of the "In/Out" decision, the next step is to assign an initial point value or unit weight to these categories. There are several different weighting combinations that may be used to score individual variables; however, there are a number of factors to be considered before assigning weights to categories of these variables. These factors include:

- the number of categories within each variable;
- the rate of incarceration within each category; and
- the desirability in terms of policy of using positive or negative scores.[28]

In this example, point values of 0, +1, and +2 have been assigned to the three categories of "prior adult felony convictions." This variable and its categories and weights would then comprise a possible factor for inclusion on the guideline scoresheet as one offender factor (see Table VI).

Variables which do not discriminate should be excluded from further analysis unless the Advisory Board requests their inclusion.

4. Development and Evaluation of Alternative Guideline Models

The project staff in each site should now begin the construction of alternative guideline models using various offense and offender scales constructed from the sentencing factors and their Burgess weights. The models and number of grids for each model will reflect the results of earlier analyses and prior policy decisions of the Advisory Board. For example, the project staff might develop two versions of the same general or statutory model, one which included arrest records and one

---

[27] See Gottfredson et al., pp. 199-206. This book provides additional references describing the use of these methods.

[28] For a more detailed treatment of the Burgess weighting system and possible point configurations, see Gelman, et al., pp. 18-20.

TABLE VI

GUIDELINE SCORESHEET

OFFENDER _____ DOCKET NUMBER _____

JUDGE _____ DATE _____

OFFENSE(S) CONVICTED OF: _____

_____

_____

OFFENSE CLASS (MOST SERIOUS OFFENSE)

[ ] OFFENSE CLASS

OFFENSE SCORE

A. Intra-Class Rank _____+

B. Seriousness Modifier _____+

0 = No injury 0 = No Weapon 0 = No sale of drugs
1 = Injury 1 = Weapon 1 = Sale of drugs
2 = Death

C. Victim Modifier (Crime Against Person) _____=

0 = Unknown victim
-1 = Known victim

[ ] OFFENSE SCORE

OFFENDER SCORE

A. Current Legal Status _____+

0 = Not on probation, parole, escape
1 = On probation/parole, escape

B. Prior Juvenile Convictions _____+

0 = No convictions
1 = 1-3 convictions
2 = 4 or more convictions

C. Prior Adult Misdemeanor Convictions _____+

0 = No convictions
1 = 1-3 convictions
2 = 4 or more convictions

D. Prior Adult Felony Convictions _____+

0 = No convictions
1 = 1 conviction
2 = 2 or more convictions

E. Prior Adult Probation/Parole Revocations _____+

0 = None
2 = One or more revocations

F. Prior Adult Incarcerations (Over 30 Days) _____=

0 = None
1 = 1 incarceration
2 = 2 or more incarcerations

[ ] OFFENDER SCORE

GUIDELINE SENTENCE _____

ACTUAL SENTENCE _____

REASONS (If actual sentence does not fall within guideline range):

Source: Gelman, et al., p. 105.

that did not. As previously stated, the intersection of the various scores from the offense and offender scales result in the creation of cells within the grids. The sentencing decisions contained in the construction sample are plotted within the cells of each grid on the basis of their score on the offender and offense scales. A FORTRAN program has been developed to perform this operation.[29] The project staff must then examine these grids containing the construction sample data to determine whether cells should be labeled "In" or "Out." The decision regarding how to label a particular cell should be made on the basis of a comparison of the number of "In" versus the number of "Out" sentences for a particular cell. The final configuration of cells should minimize, insofar as possible, errors in predicting the "In/Out" decision. Furthermore, the logic of the guideline concept is "as the offense becomes more serious, and/or the offender's unfavorable characteristics become more pronounced, the probability of incarceration...should increase."[30] Therefore, each cell must be examined with those contiguous to it to make those modifications necessary to conform to this logic.

After the cells are labeled, the models found to discriminate best in terms of the "In/Out" decision are presented by the project staff to the Advisory Board. The project staff at this time must also decide whether to include additional information concerning length of sentence for "In" cells when presenting the model to the Advisory Board.

At the Advisory Board presentation, the guideline models found to be most predictive are reviewed. The Board may reject some models, accept some outright for testing using a validation sample, or direct the project staff to modify some models. The Advisory Board may want some of the models adjusted either by incorporating variables previously rejected or not used by the project staff, or by changing the weights assigned to particular variables comprising the offender and offense scales of the models. Should modifications be necessary, the project staff should repeat the appropriate steps of the model development process.

5. Validation of Guideline Models

Once the Advisory Board in each site has selected a number of models as possible representations of their sentencing policy, the project

---

[29] Gelman, et al., pp. 160-182.

[30] Gelman, et al., p. 12.

staff should commence the testing of these alternative models. Since the models were developed and tested on one sample, social science methodology requires that the predictive accuracy of these models be tested on another sample to assess their validity. The results of the validation will be critical to the Advisory Board in their final selection of a guideline model. Each model will be tested using a second or validation sample of sentencing decisions. The grids should be validated both within and across participating jurisdictions of a site in order to examine the fit of the models. (Project staff may also want to employ this procedure at other points in the developmental process, for example, with the construction sample.) The data collection instruments used to gather this sample will closely resemble those developed for the construction sample. They should incorporate improvements in the specification of variables as a result of the staff's experience with the construction sample. The data collection instrument for the validation sample will focus primarily on those variables incorporated in the alternative models and exclude as far as possible variables not used in these models. The procedures followed in collecting the data for model testing are the same as those used with the construction sample (Section II C).

The validation sample should be selected from the most current pool of sentencing decisions available. The time frame of the validation sample will be closer to the point of implementation of sentencing guidelines than the construction sample. The collection of the validation sample should commence as soon as possible after completing the analysis of the construction sample and the presentation of the results to the Advisory Board. Previous site experience suggests that the validation sample should be approximately one-half to two-thirds the size of the construction sample.[31] The validation sample should be drawn in the same fashion as the construction sample (Section II C).

Once the data have been prepared for analysis, the project staff should begin evaluating the models selected by the Advisory Board. The predictive accuracy[32] of the models may be tested by applying the computer program used to develop the preliminary guideline models (based on the construction sample) to the validation sample. The validation sample may also serve another purpose. It can be combined

---

[31] See Gelman, et al., p. 21.

[32] This refers to the "In/Out" decision. With the assistance of the project staff, the Advisory Board may devise one or several approaches for testing the guidelines' accuracy for predicting sentence length.

with the construction sample to provide additional experiential data within the guideline models. The Advisory Board may find this desirable as it increases the number of cases within each cell on which decisions are based.

6. Selection of a Final Guideline Model

As a result of this validation process, the project staff should be able to identify which models continue to be the most predictive of sentencing decisions. These results will be presented to the Advisory Board for their consideration so they can select one model to represent their sentencing policy. Once they have selected a model, they may wish to make adjustments to it (e.g., reconciling inconsistencies regarding length of incarceration among adjacent cells).

At this meeting, the Advisory Board should also decide the final format of the guideline model they have selected. There is a variety of information items related to sentencing decisions (as well as methods of displaying these items) which can be presented within the grids. These items include:

- the labeling of a cell as "In" or "Out";
- some presentation of the ratio of the "In" to "Out" decisions within each cell;
- some form of tabulating actual sentencing decisions;
- the range of sentence length within each cell; and/or
- some measure of the central tendency of sentences within cells.

Although the final decision as to the contents of the cells is the responsibility of the Advisory Board, at a minimum, the cells of the sentencing grids should contain suggested sentences in terms of the "In/Out" decision and length of incarceration. It is recommended that the prescribed sentence range represents only a portion of the range of sentences contained within each "In" cell. For example, it may be decided to use the range represented by the central 50 percent of the cases sentenced.

E. Model Implementation

Until now, this test design has been concerned with the development of a set of grids designed to articulate the sentencing policy of judges in each of the sites. In order to aid in the reduction of

sentencing disparity, the guidelines must be used by judges on a consistent basis. To ensure this routinization, the guidelines must be formally implemented and a support mechanism developed to facilitate the daily operation, review and institutionalization of the guidelines; analyze the sentencing data; conduct review sessions for the judges; and modify the guidelines as directed.

Once the Advisory Board in each site has agreed upon the adoption of a specific set of sentencing guidelines, this group should make an appropriate presentation to the judges in the participating jurisdictions of each site. The purpose of this presentation is essentially informational, that is, to introduce the guidelines and their supporting documentation to the judges. This documentation will primarily consist of:

- a coding manual;
- the guideline scoresheet; and
- the guideline grids.

The coding manual will specify and define the variables used in the sentencing guideline model. This manual provides the basis for translating the information contained in case files into the offense and offender scores which comprise the axes of the sentencing grids to be used. For example, if the variable "Criminal Status at the Time of the Instant Offense" was used in the offender score, it might be defined in the following manner:

<u>Criminal Status</u>

0 - Free
1 - Not Free

<u>Code "0"</u> if the offender was not under any form of criminal justice control. Include here if in the military or voluntary hospitalization (e.g., <u>not</u> court-ordered hospitalization).

<u>Code "1"</u> if at the time the offense was committed the offender was under any form of state control as a result of some civil or criminal action (e.g., AWOL, sex offender). This includes the offender with any charges, adult or juvenile, which have not yet been disposed of, including those persons on pre-trial release or awaiting sentencing on bail, bond, ROR, or in jail. Code "1" includes persons on deferred prosecutions, deferred judgment, conditional release, outstanding warrants. It also includes those offenders incarcerated at the time of the offense and those on escape status.

The actual score for this variable in each particular case would be recorded on a sentencing guideline scoresheet (see Table VI). A similar procedure is followed for scoring all of the other items included in the guidelines. As indicated in Table VI, separate scores are calculated for the offense and offender scales. These scores are then located on the appropriate guideline grid. The intersection of offense and offender scores locates a specific cell, within the appropriate guideline grid, containing a suggested sentence.

To ensure the uniform application of sentencing guidelines by the judges within each site, the guidelines should be formally implemented by court rule, by decree of the Chief Justice of the Supreme Court or by formal agreement among the participating judges. For this test design, formal implementation means that the judges will in each case:

- consider the guideline sentence;
- record the sentence actually imposed; and
- provided written reasons when departing from the suggested sentence.

A completed guideline scoresheet and the appropriate grid indicating the guideline sentence must be presented to the judges prior to sentencing. The guideline scoresheet may be filled out by the judges, members of their staff, or probation officers. In any event, experience indicates that computation of the scoresheets should be assigned to one organization which would have continuing responsibility for this task. Whoever is assigned this responsibility must be trained by project staff in the methods for completing the guideline scoresheet.

Concurrently, the judges must also receive information from the project staff concerning the use of the guidelines. The judges need to understand the importance of articulating specific reasons when they impose a sentence other than that suggested by the guidelines. For example, the guideline model used in a particular site might not incorporate the variable "employment history." Yet in some cases, a judge might cite a history of good employment as his reason for imposing a less severe sentence than that suggested by the guidelines.

The project staff should devise a system to ensure that the scoresheets are distributed to the organization responsible for calculating the guideline sentences and forwarding the scoresheets to the judges. The staff should also develop a mechanism for ensuring that the completed scoresheets are collected and returned to them for monitoring, analysis and review. While the individuals who prepare the guideline scoresheets are responsible for their accuracy, project staff must

continually monitor the sheets for accuracy and completeness. Working with the Advisory Board, the project staff should provide whatever continued education and information are indicated by their monitoring efforts. In addition to monitoring the accuracy and completeness of the sentencing guideline scoresheets, the staff must also analyze the information contained on these sheets to:

- determine the extent to which sentencing decisions are falling within the guidelines; and
- indicate where the guidelines might require modification.

The first step in this process is the preparation of the data contained on the scoresheets. A coding manual must be prepared which translates the information contained on these sheets into a format which can be stored for later computer-assisted analysis.[33] Once the computerized data base has been developed, the information must be checked for errors using the same process as applied to the construction and validation samples.

When the accuracy of the information has been checked, two types of analysis should be performed:

- an examination of sentencing variation within the cells of the guideline grids; and
- an examination of those cases falling outside of the guidelines.

The first type of analysis will examine the extent to which guidelines have affected sentencing consistency. Essential to this analysis is a detailed examination of the percentage of cases that fall within the appropriate cell of each grid and of changes in sentencing variation.

The second type of analysis will examine those sentences which differed from the guideline sentence. The purpose of this task is to identify patterns indicative of the need to modify the guidelines and, on the basis of these patterns, to suggest appropriate changes. The first step in this analysis is to specify and categorize by reason all cases where the sentences differed from those suggested by the guidelines. The project staff must then examine each category to determine what additional information (not already included in the guideline system) had been used by judges in these cases. For

[33] Gelman, et al., pp. 183-198.

example, the analysis may reveal that the judges have been frequently imposing harsher sentences than suggested by the guidelines in assault cases, citing as their reason that the crime was against an elderly person. The existing guidelines must then be reviewed by the project staff, in light of the additional information being considered by the judges and the actual sentence imposed, to determine if changes in the guidelines are indicated. As a result of this review, the project staff should then be in a position to recommend any appropriate changes in the guidelines. The changes may be of three types--in the weight of variables already in the guidelines; in the addition of new variables to the guidelines; or, in the adjustment of the type or length of sentence. For instance, in the example given above, it might be recommended that the offense scale be adjusted to provide additional weight when there is a crime against the elderly.

This test design also requires that periodic review sessions be conducted with each site's Advisory Board. The first of these review sessions should be held not later than four months after the implementation of the guidelines. The purpose of these sessions is to review, on at least a semiannual basis, the performance of the guidelines and to determine what, if any, modifications are to be made to the guidelines as a result of the analysis conducted by the project staff. Additionally, the review sessions provide the opportunity to consider policy issues related to sentencing beyond those based on the analysis of past sentencing practices. The importance of the review sessions to the successful use of sentencing guidelines should not be underestimated. The willingness and ability of each site's Advisory Board to review and modify the guidelines will contribute powerfully to the viability of the guidelines as a dynamic representation of sentencing policy.

III. EVALUATION

A. Introduction

The purpose of this section is to identify the objectives of this test and some of the evaluation activities considered necessary for their examination. Additional information on the evaluation effort is set forth in the NILECJ solicitation for the evaluation of this test. An independent organization will be chosen by the Institute to conduct an evaluation of each of the sites selected to develop and implement multijurisdictional sentencing guidelines. The major objectives of the evaluation are:

- to test the effectiveness of guidelines as a method for increasing sentencing consistency;
- to assess the feasibility of developing and implementing sentencing guidelines as a policy tool in a multijurisdictional setting;
- to examine the impact of guidelines on sentencing practices and on other components of the criminal justice system; and
- to provide a descriptive account of the guideline development and implementation process for use by other jurisdictions interested in this process.

These four objectives address both the outcomes (Objectives One and Three, above) and processes (Objectives Two and Four, above) of the project. The evaluator will be expected to work closely with project staff in order to collect the qualitative and quantitative data needed to address these objectives. The evaluation period will be 2-1/2 years, beginning with the start of the project in each site.

The evaluation approach outlined here is designed to produce not only knowledge of the impact of guidelines for the evaluation/research community and for jurisdictions considering the development of guidelines, but also technical descriptions of the guideline development process for use by those undertaking this process. The analytical approaches described are in no way definitive or exhaustive of the possible methodologies and data which might be fruitfully employed to address these objectives.

B. Evaluation Objectives

1. To Test the Effectiveness of Guidelines as a Method for Increasing Sentencing Consistency

The proposed design for assessing the impact of guidelines on sentencing makes use of before-after comparisons that examine changes in a variety of measures from a baseline period (before guideline use) to two "program" periods during which guidelines were used. The use of a second "program" period is recommended because this second period can be employed to analytically represent those cases sentenced after the first review and modification of the guidelines. Measurements during this period may reflect the impact of revised guidelines.

At a minimum the evaluator should use the sentencing measures discussed below in making outcome assessments. These should include changes in the following measures of central tendency and dispersion (calculated for each cell of the guideline grid):

- percentage of offenders incarcerated and not incarcerated;
- percentage of decisions within the suggested sentence range;
- mean sentence for those incarcerated;
- median sentence for those incarcerated;
- sentence range for those incarcerated; and
- sentence variance for those incarcerated.

One of the major objectives of the evaluation is to test the effectiveness of guidelines as a method of improving sentencing consistency (or conversely, reducing disparity). One conception of disparity is based on the notion of "unwarranted" variation in sentencing. This definition necessarily involves normative judgments regarding the validity of the sentence and factors used in determining it. This evaluation, however, requires that some quantified measure of consistency/disparity be developed and that this measure be as objective as possible (that is, not judgmental). This test is based on the expectation that whatever the ratio of warranted to unwarranted sentencing variation may be for similar offense/offender types, the total amount of variation should decrease if guidelines are effective. As noted in Section II A of this document, sites will be required to identify the level of variation in existence prior to guideline implementation.

Although the statistical variance in sentences could be employed as a measure of consistency/disparity, this would not reflect differences in the percentages of offenders incarcerated and not incarcerated. Thus, a continuous scale of sentence severity should be employed which quantifies all possible sentencing outcomes including probation, split sentences, and so on. For example, Diamond and Zeizel[34] have adapted a sentence severity scale of this type for use in analyzing disparity. By assigning scale values to all sentences, disparity is expressed as the average difference in sentence between two judges sentencing similar offenders, expressed as a percentage of the mean sentence. Whatever scale is developed, it should be examined in terms of local judicial perceptions of the relative severity of various sentences.

This percentage disparity measure should be calculated for all cells of the sentencing grid. The central hypothesis to be tested, then, is that guidelines will reduce the sentencing differences for offenders calculated to be within the same cell. The criteria established by the judges for the reduction of disparity will be used to assess the significance of changes from the baseline period (phase I) to subsequent program periods (phase II & III).

A corollary approach to testing the effectiveness of guidelines as a means of improving both inter- and intra-cell consistency involves the use of multiple regression to predict sentencing decisions during the three analytical periods. By coding all cases with respect to the jurisidction and individual judge involved, it is possible to assess (via multiple regression) the systematic contribution of these two factors to sentencing decisions. It is hypothesized that the effect of sentencing guidelines should be to reduce (from baseline to program periods) the influence of individual judges and jurisdictions on the determination of a sentence. It should be noted that coding cases by jurisdiction is essential to allow a clear description of differences in jurisdictional sentencing practices.

2. <u>To Assess the Feasibility of Developing and Implementing Multijurisdictional Guidelines</u>

The assessment of the feasibility of a multijurisdictional guideline approach represents the major process objective of the evaluation. As such, this assessment asks whether the development and implementation

---

[34] Shari Seidman Diamond and Hans Zeisel, "Sentencing Councils: A Study of Sentence Disparity and its Reduction," <u>University of Chicago Law Review</u>, V. 43, 1976, pp. 109-149.

of multijurisdictional guidelines can be accomplished within the parameters outlined in this test design. In order to address the feasibility issue the evaluator will have to collect a range of qualitative and quantitative information bearing on the development and implementation processes.

First, the evaluator will be expected to monitor and collect information descriptive of the successive stages of guideline development, implementation, and revision. Distinct activities such as variable selection, data collection, model selection and development, and model implementation and revision should be examined as logical, sequential activities to be undertaken and completed collaboratively by the project staff and the Advisory Board. Additionally, the evaluator should examine the extent to which guidelines are serving as a mechanism for the articulation of an explicit sentencing policy. The evaluator should attend all Advisory Board meetings so that the policy inputs/decisions of the judges during guideline development and review can be identified and described. Additionally, the way in which these inputs are represented in the guidelines should be described to allow the assessment of whether changes in sentencing measures are related to these explicit policy decisions.

The feasibility assessment should also examine the use of guidelines by the judges. Although an Advisory Board of judges will be formed in each site to represent the total body of judges, the evaluator should develop survey instruments so that periodic assessments of the attitudes and perceptions of all judges using the guidelines can be made. This will allow the description of the judges' perceptions of guidelines and their purposes; of their attitudes toward the utility and influence of guidelines; and of their methods in using or considering the guidelines. Additionally, the evaluator should collect data from the guideline scoresheets (see Section II E) to measure the extent to which the judges are recording their actual sentences and providing reasons for sentences falling outside the guidelines. In effect, this addresses the question of formal implementation.

3. <u>To Examine the Impact of Guidelines on Sentencing Practices and Other Components of the Criminal Justice System</u>

The evaluator should examine the awareness, role, and influence of other CJS agencies and actors who are directly or tangentially involved in the sentencing process. There are a number of techniques which can be employed to investigate the extent to which prosecutors, defense attorneys, corrections agencies, and others are aware of guidelines and whether their rules/behavior have changed or adjusted because of guideline use. For example, it will be important to know whether decisionmaking at the screening and charging stage has

changed, whether diversion or plea bargaining practices have been affected, and whether "judgeshopping" or other defense strategies have been influenced. In any case, the evaluator should consider the range of system effects, or unintended effects, which might occur and devise appropriate research strategies for exploring these effects.

4. To Provide a Descriptive Account of the Guideline Development and Implementation Process for Use by Other Jurisdictions Interested in the Process

The evaluator will be expected to provide extensive historical descriptions of the successive stages of guideline development, implementation, and revision. The various distinct activities of these stages (e.g., variable selection) should be examined as decision-making processes involving tradeoffs between a variety of methodological, policy, legal, and political concerns. Especially critical will be the development of detailed descriptions of technical aspects of data collection and model development including the statistical outputs generated in these stages. In this way, the evaluators will have to develop an extremely close working relationship with the project staff so that they accurately document the day-to-day choices, alternatives and conflicts that arise and the methods of resolution that are selected.

C. Some Evaluation Considerations

It should be noted that the qualitative and quantitative information collected in order to address one objective of the test may prove useful, or even essential, to the assessment of another objective. For example, although these statistical analyses of sentencing and court processing measures (Objectives One and Three) will indicate what changes have occurred and whether these changes are significant, they do not, in and of themselves, allow the attribution of these changes to guidelines. The use of simple before-after analyses, without control groups, allows the possibility that a wide range of forces and factors, including the guidelines, may be causing these changes. For this reason, the evaluator must use a wide range of information from the process assessment in conjunction with the outcome assessment in order to achieve a reasonable basis for attribution. It will be necessary to consider a range of factors influencing sentencing including:

- explicit policy decisions of the judges;
- the prescriptive pressure of the guidelines;

- changes in prosecutorial or defense behavior; and
- legislative and public pressures.

A significant aspect of the evaluation is that much of the data required for the assessment of the changes in sentencing and sentencing variation will be the same as the data collected as part of guideline development. Thus, it is critical that the evaluator coordinate data needs (including sample considerations) with the guideline developers from project start-up through implementation. In this way unnecessary duplication of data collection efforts can be avoided and the evaluator can ensure that special data needs related to the evaluation can be integrated into the regular data collection processes.

IV. IMPLEMENTATION AND NILECJ SUPPORT

A. Implementation

The proposed test effort has been designed for implementation within several jurisdictions of each of two or three states (each state must select only one urban jurisdiction, at least one surburban and, if available and supportable, a rural jurisdiction). This distribution of sub-site participation will provide the diverse environments required for this field test. The test is designed in three stages over a twenty-four month period (see Figure 3, p. 10, for a detailed description of stages and specific tasks that comprise each stage). The initial stage will involve up to seven months of data collection, training of program personnel and strategic planning.

The second stage will involve about five months of model development, assessment and selection. The specific tasks to be conducted during this period relate to the construction and validation of alternative guideline models and the selection of a specific implementation model.

One of the outcomes of this stage will be the identification by each site's central staff of the existing levels of sentencing variation. This identification of variation must be accomplished within the first 12 months from the start of the program. The sub-sites are responsible for cooperating in the collection and analysis of the data that will be performed by the central staff to determine the existing levels of sentencing variation within their jurisdiction. The identification of baseline sentencing variation will then be used as a basis for determining the success of the test effort (see pp. 10-11).

The final stage will cover twelve months and involve the implementation and periodic review/revision of the guidelines within each jurisdiction by all affected judges. In addition, the site will be responsible for cooperating with the national evaluator in measuring the impact of the guidelines on the level of variation identified in Stage Two.

B. NILECJ Support

NILECJ support will be provided in the form of financial assistance and training. A consulting firm will be retained by the Institute to provide implementation assistance to the participating

jurisdictions. Support will include training for key program personnel, consultant services to aid program sites in the planning and implementation of the program elements to be tested, and various conferences and meetings to enable key personnel from each of the participating programs to discuss problems and issues of mutual concern. Funds will also be included to support research utilization efforts such as hosting visiting court officials so they may observe program operations.

NILECJ will allocate approximately $300,000 per state for participation in the program. The funds will cover the cost of a project director, research director, research analyst, and computer programmer and the associated expenses for data collection, processing and program operations by the central staff. No funds for hardware will be provided by NILECJ.

The grantee must provide reasons for recommending sub-sites to participate, using the site selection criteria outlined in Section V of this test design document. The final selection of subgrant sites will be made by the NILECJ. The central staffing patterns must be designed in a manner most consistent with the tasks outlined in Figure 3 (p. 10) of this test design document.

C. Implementation Definitions

To assist grantees in the development of their program plan it is important that the following terms be understood:

1. A site is composed of the participating sub-sites and the central staff of the project;

2. The sub-sites are the local court participants in each state;

3. Central staff are project staff responsible for coordinating the program tasks among all the sub-sites and assisting the sub-sites in the implementation of the program;

4. The grantee should be the agency responsible for the superintendence and/or administration of sentencing policy within the state;

5. The term cases should be interpreted as referring to sentencing decisions by the courts of general jurisdiction in each participating jurisdiction;

6. The data base is the sample of sentencing decisions used to develop the sentencing guideline instrument.

V. SITE SELECTION

The site selection criteria for the field test design of multijurisdictional sentencing guidelines were initially developed by NILECJ's Test Design Group based on previous guidelines experience; the objectives of the test design; and consultations with state court administrators, judges and other court personnel during the latter stages of design development. As indicated throughout this document, NILECJ plans to implement the test in the courts of several jurisdictions within each of two or three states. Each site shall contain no less than three nor more than five sub-sites. Only one of the participating jurisdictions within each state can be urban. Jurisdictions may be defined differently from state to state depending on the nature of a state's judicial system and the congruence of its divisions with political and/or geographical units. Participating courts must be courts of general jurisdiction having original jurisdiction over felony type offenses.

The site selection criteria are divided into two categories. The first category consists of those criteria which are considered essential for the successful development and implementation of sentencing guidelines as detailed in this test design. The second category is composed of those criteria which, while not considered essential, would nevertheless facilitate the development and implementation of guidelines. Final selection of test sites will be made by NILECJ based on an assessment of the relative strengths of each applicant in terms of all criteria.

A. Criteria Considered Essential to Guideline Development and Implementation

The following criteria are considered essential to the development and implementation of sentencing guidelines.

1. Commitment by the site that the development and implementation of guidelines will follow the test design detailed in this document.

2. Absence of current or pending legislation or ongoing sentencing activities (such as some form of mandatory sentencing) which would limit discretion and/or obviate the purposes of this test.

3. Indication of interest, cooperation and commitment on the part of the judges and other court personnel who would participate in the test. A resolution passed by the judges to be involved is one possible method of satisfying this criterion.

4. Commitment by the site that during the two year test period the implementation of guidelines would be formally mandated. In this context, the term "implemented" means that once a guideline model has been formally adopted (see pp. 9-11, 36-37), the judges participating in the test will consult or use the guidelines in conjunction with all their sentencing decisions, that is, in each sentencing decision, the judges will:

   - consider the appropriateness of the sentence suggested by the guidelines in terms of the calculated offense and offender scores for that case in view of all the other information available at sentencing;

   - record the sentence actually imposed; and

   - provide written reasons specifying why the guideline sentence is inappropriate in any particular case.

   There are a variety of means by which guidelines, once adopted by a site, might be formally mandated, e.g., through rule or administrative directive of the Supreme Court or through a formal agreement of all the judges participating in the test. The most appropriate means of formally mandating the implementation of sentencing guidelines may vary from state to state.

5. A total annual sum of the sentencing decisions (cases) in the courts proposed for participation of at least 4,000, with no more than three-fourths and no less than one-fourth of all cases originating from any one jurisdiction. A sentencing decision should be construed as the sentence imposed on any one offender for any one crime or series of related crimes. For purposes of this test, participating courts must be courts of general jurisdictions, or courts of the first instance, which have original jurisdiction over felony type offenses.

6. Commitment by a site to cooperate with the national level evaluation of the development and implementation of multi-jurisdictional sentencing guidelines as specified in this test design.

7. The availability and retrievability of case information from each participating court recording the information available to the judges at sentencing. Examples of such documents are presentence investigation reports, criminal history records and descriptions of the offense(s) involved.

B. Criteria Facilitating Guideline Development and Implementation

The following criteria facilitate the development and implementation of sentencing guidelines.

1. Some prior recognition of sentencing variation and/or the need to provide judges with a tool to assist them in making sentencing decisions. This recognition might be evidenced in many ways, e.g., through empirical studies or via sentencing institutes or judicial study groups designated to address the problem of sentencing variation.

2. The project director be available and, if possible, assigned from the agency responsible for the superintendence and/or administration of sentencing policy in the state.

RESOURCE DOCUMENTS

The American Friends Service Committee. Struggle for Justice: Report on Crime and Punishment in America. New York: Hill and Wang, 1971.

Calpin, Joseph C., Jack M. Kress, Marilyn A. Chandler, Arthur M. Gelman and Barbara A. Broderick. The Analytical Basis for the Formulation of Sentencing Policy. Unpublished, January 1978.

Fogel, David. We Are the Living Proof, The Justice Model for Corrections. Cincinnati: W. H. Anderson, 1975.

Frankel, Marvin. Criminal Sentences: Law Without Order. New York: Hill and Wang, 1972.

Gelman, Arthur A., Jack M. Kress and Joseph C. Calpin. Establishing a Sentencing Guidelines System: A Methods Manual. Washington, D.C.: November 1977.

Gottfredson, Don M., Leslie T. Wilkins, and Peter B. Hoffman. Guidelines for Parole and Sentencing: A Policy Control Method. Lexington, Massachusetts: Lexington Books, 1978.

New York State Special Commission on Attica. Attica: The Official Report of the New York State Special Commission on Attica. New York: Bantam Books, 1972.

The Twentieth Century Fund Task Force on Criminal Sentencing. Fair and Certain Punishment. New York: McGraw-Hill, 1975.

Von Hirsch, Andrew. Doing Justice: The Choice of Punishment. New York: Hill and Wang, 1975.

Wilkins, Leslie T., Jack M. Kress, Don M. Gottfredson, Joseph C. Calpin, and Arthur M. Gelman. Sentencing Guidelines: Structuring Judicial Discretion. Washington, D.C.: February 1978.

U.S. GOVERNMENT PRINTING OFFICE: 1979 281-380/4539 1-3

END