A SAS BASED MANAGEMENT INFORMATION DATA BASE

Wayne B. Finley and William C. Pannell, California Board of Prison Terms

Introduction

Under the provisions of the Uniform Determinate Sentencing Act of 1976, the California Board of Prison Terms is required to review the sentences of a majority of the offenders sentenced to state prison. The Board is required to review the sentences of persons received in prison with determinate sentences to determine if the sentences are substantially different from those imposed on other offenders convicted of similar crimes under similar circumstances.

The wording of the legislative mandate to review sentences offered little hint of how sentences should actually be reviewed. The extent of thinking in 1977 was that a great number of factors might have something to do with sentencing. Meetings were convened at that time to discuss the conceptual aspects of disparate sentence review. These meetings included the Board of Prison Terms (then the Community Release Board), the Department of Corrections and the Rand Corporation. Other consultants were also sounded by the Board.

An extensive list of potentially useful data items was developed. Most of the originally suggested variables exist in the Board's data base today. A practical conceptionalization of the review mandate was also developed.

It was determined that the review would be one of judicial discretion, not, for example, widely discretionary prosecutional discretion. It was felt, although not unanimously, that the mandate was to review the sentences of persons "convicted of similar crimes .....".

It was clear that some sort of computerized statistical approach was appropriate, although exactly what approach was not clear. Among the suggestions were regression analysis to predict sentences, log linear modeling, numerical taxonomy to group offenders, factor analysis to group variables, discriminant analysis and various combinations of the above.

The Board soon realized that in-house technical staff and computer capability was a necessity.

Staff Organization

When the Board, in compliance with Penal Code Section 1170(f), organized the Data Processing Unit and Management Information Section, a conscious effort was made to minimize the traditional barriers of communication between research and data processing. In the State of California many research and data processing projects have failed due to organizational barriers between data processors and researchers.

To prevent those traditional barriers from forming, the following policies were developed:

1) Both units occupy the same office space
2) Researchers are free and expected to design and write their own computer programs
3) All computer resources in the office are shared, i.e., access to terminals, printers, etc.
4) Data processing staff act as consultants on technical DP problems and promote access to terminals and the use of good programming practices
5) No policy barriers to communication would be established, i.e., "a MIS staff member can talk only to the Chief of data processing for consultation"

The results of these policies were gratifying. Staff exchanged ideas freely on the design of the system. Many informal structured walk-throughs were conducted without staff even knowing they were participating in the "latest method" to build a quality computer system. DP and MIS staff worked as an integrated team. Each member brought expertise to achieve the common goal of building a computer system to review prison sentences. There were plenty of design disagreements and arguments, but due to the office atmosphere most were resolved through frank and open discussion of the problems.

With adoption of SAS as the main programming language, policy 2) was extremely easy to implement. We had the good fortune of having the Data Center lease SAS just as the project began. It was extremely easy to teach researchers to write good program code without an inordinately long learning curve. All research staff have learned to write SAS code without a great amount of trauma. The code is surprisingly readable, not like reading a Fortran program written by a novice programmer.

A natural division of labor occurred between staff. The DP group became responsible for data entry, editing, correcting and updating the data base. They also are responsible for daily operation of the local computer system and normal operation of the review process. The MIS group, the research staff, concentrated on the design and implementation of the main project goal, sentence review of felon prison cases. The staff freely exchanged ideas on the design of the data base, the interfaces between data entry, SAS data base, and problem solution. DP staff handled JCL and data management problems. The data content problems were fielded by MIS. The project size, no doubt, aided the free interchange of ideas. Larger projects have a more difficult time in developing a team atmosphere.

Project Team

The size of the Board of Prison Terms, only 100 staff, dictates that a DP unit and MIS unit be kept small. However, due to the complexity of

sentence review and the statewide impact, the positions were established at a higher level than normal. All original staff of the team had several working years in either data processing or research projects. At the present time, the total number of personnel is nine. The positions are as follows:

DP Staff
Chief, Senior Programmer Analyst
Associate Programmer Analyst
Data Processing Technician
Key Data Operator

MIS Staff
Chief, Operations Research Specialist III
(3) Associate Research Analyst
Secretary

Computer Equipment

The Board leases a IV-Phase IV-90 computer for its in-house computer system which connects, via direct telephone lines, to the Teale Data Center, a state-owned multi-processor facility. The local system contains the following:

Size
484k bytes main memory
67 mega bytes disc space

Devices
18 video display units
300 LPM printer
(3) 55cps printers

Communications
RJE line to the Data Center
3270 line to the Data Center

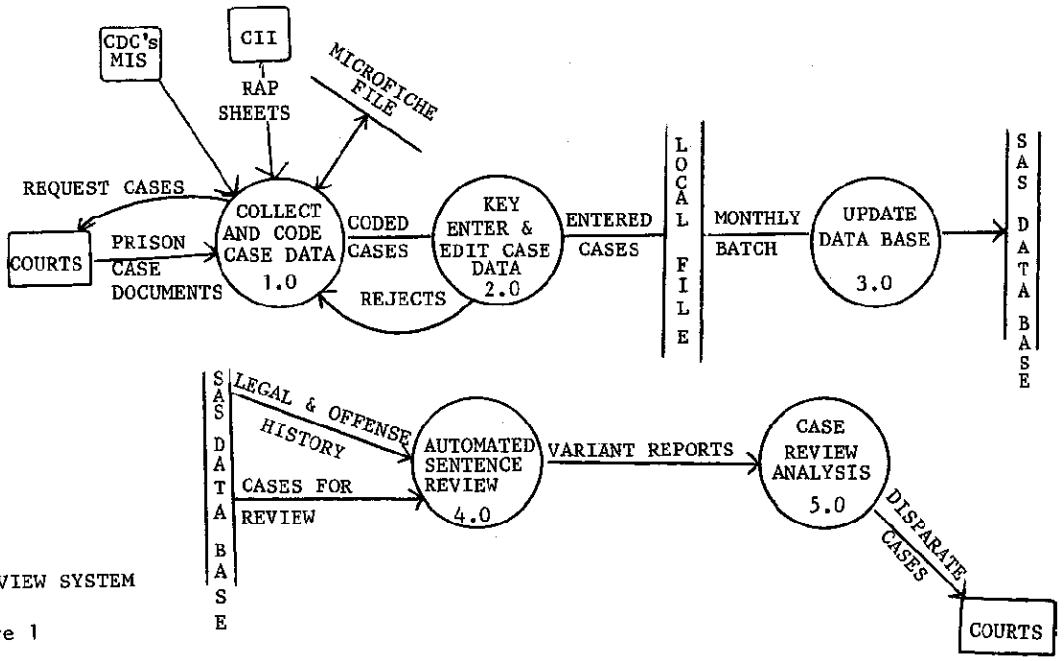
The local system is used for key entry, word processing, program development, remote job entry, TSO and CICS on-line sessions with Teale Data Center. The key entry facility allows one to write highly extensive format programs to edit the data while it is being key entered. It also is an excellent tool for program development. All SAS code and JCL code reside locally on the disc. Each programmer and research analyst has his own video terminal. The system is similar to a programmer work station.

The SAS data base resides at the Data Center. To execute a SAS program, the RJE processor is invoked to transmit the JCL and SAS code. Output can be brought back as print or punch to be stored on the local disc file. The punch data can then be processed locally for interactive key entry or a word processing application. A typical use for key entry would be to use punch data as identifiers stored on a local index set. The key entry process then could validate the identifiers when the key data operator enters the new data. For word processing, the punch file could be a name and address list. The list is merged with a form letter to produce a typewritten addressed letter.

The Sentence Review System Process

Figure 1 is a data flow diagram of the sentence review system. Each circle represents a process of the system. The lines going into and out of the circles represent the data inputs and outputs. The boxes are the data sources and sinks. The | | are data files.

The five basic processes of the system are data collection and coding, data edit and entry, SAS data base update, automated sentence review, and case review analysis. Of interest for this paper are data edits and entry, SAS data file structure, and the Automated Sentence Review.



SENTENCE REVIEW SYSTEM

Figure 1

Data Entry

For each case coded, a minimum of 172 data elements are collected. The total amount of data coded per case will vary with the number of charges, victims, and legal convictions. A case could have several coded pages to enter. About 1200 to 1300 cases are received each month.

Data elements are composed of fixed length profile information, and variable data for charges, victims, and legal convictions. The profile segments are juvenile and adult criminal history, personnel information, and offense information.

The data entry function is not a standard key punch process. The large number of data elements per case and the variable nature of the data does not lend itself to simple 80 column records. With the IV-Phase key entry facility, one can design multi-screen key entry formats. Each screen represents a fixed-size record. An entered prison case is composed of 12 different sized fixed length records. For the variable data, there are mutiple records. For any case the total number of records entered is variable and a function of the number of charges, victims, and legal convictions.

The key entry tool also allows extensive editing to be done while the data is being entered. We have defined two edit types, validation and consistency. A validation edit checks for range and data type. A consistency edit compares one data element to another data element for consistency. For example:

"The month of a date field must be 1 through 12" is a validation edit.

"The date of birth must be less than the date of first adult conviction" is a consistency edit.

The long delay of a batch editing process is eliminated. Errors can be found and corrected soon after coding. This interactive data entry and editing behaves like an on-line CICS application without the poor response time and the high cost in hardware and personnel at the Data Center.

The results of this early extensive editing allow a high degree of confidence in the accuracy of the data. We believe the sentence review data base to be one of the "cleanest" in the criminal justice field.

SAS Data Files Structure

Each month we transmit a monthly batch to the Data Center to update the SAS data base. Each case in its raw entered data form is a collection of a variable number of sequential records of different sizes. To handle the variable, or periodic data problem, on advice from a SAS instructor, we designed a relational data model for our SAS files. Figure 2 represents the data file structure. All fixed information is stored on a master file. The variable data is stored on subordinate files. There are multiple records per case on each subordinate file. Each record has unique identifiers to define it as a member of a case and an identifier which separates it from other case records.

To transform the data from key entered records to SAS records, a local batch program decomposes the monthly batch file into a master file and subordinate files. These temporary files are transmitted to the Data Center as card images (the IBM system thinks all RJE's are card readers). A SAS program inputs these files, calculates summary variables from the subfiles, and puts those summary variables in the monthly master file. It then adds that month's data to the permanent SAS files. Upon completion of the update, the monthly batch is then processed by Automated Sentence Review.

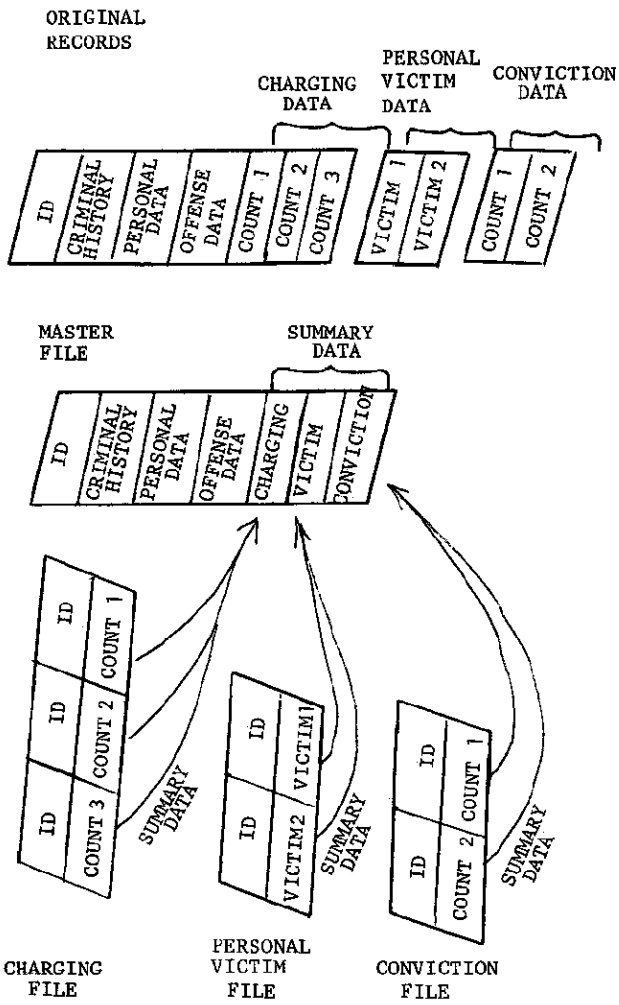


Figure 2

Sentence Review

It was decided that the Board would attempt to develop several different methods of sentence review to be used in "series" with one another.

The first type of automated review to be developed was a Monte Carlo simulation model. This model attempts to replicate the decisions that a judge makes when sentencing.

The model was implemented in PLI, a high level IBM language. The program conforms to most standard structured modeling concepts. The program is top-down modular in design and the code is rich in comment and lacking in go-tos. It is a complex program that is understood and used by more than one staff member.

Input For Simulation Model

The Automated Sentence Review (ASR) model reads two types of data. First it reads inmate specific data, describing a prisoner to be reviewed. This information includes the detailed legal circumstances of the offenses as well as information such as injury inflicted by the offender to victims, criminal justice status of the offender at the time of his conviction offense and the prior prison experience of the offender. The second type of data used by the model consists of odds or probability distributions that relate to the several sentencing decisions. These probability distributions are derived from observed sentencing practices and address the several decision components of the Determinate Sentencing Law.

The Model

Sentencing under the Uniform Determinate Sentencing Law of California is a precisely defined process. There are three possible "base" sentences for any particular offense. This base sentence can be enhanced or increased in several ways. Enhancements include increasing the sentence for use of a firearm, great bodily injury, and prior prison terms. Time may also be added for offenses in addition to the primary offense. The law also provides complex limitations to sentencing under certain circumstances.

The model is programmed in PLI to replicate the logic and special limitations of the law. The model calculates sentences for every offender received in prison under the Determinate Sentence Law, then compares the calculated sentences to the actual sentence received by that person. Sentences are calculated on the basis of the particular conviction circumstances of the offender and the observed historical sentencing patterns or probabilities drawn from the data base.

A judge sentencing an offender is typically faced with many choices. His decisions compound to produce a sentence. Considerable variation is possible in the sentence finally selected.

The simulation program (ASR) calculates a sentence for each prisoner using Monte Carlo techniques. Sentences for each offender are calculated 10,000 times. If, according to the simulation, fewer than 10.5% of those 10,000 theoretical sentencings would have resulted in a sentence as high or higher than the actual sentence imposed in that case, the case is identified as requiring further review. A similar process identifies cases which may be disparately low. Cases identified by the ASR for further scrutiny are submitted to secondary screening by a Hearing Representative and other Board staff.

Output

The output of ASR consists of an OS file each record of which contains a distribution of 10,000 sentences for each offender. Additionally, the records contain statistics describing each distribution. These statistics include several measures of central tendency such as the mean, several measures of dispersion such as the standard deviation, the most likely sentences and the least likely sentences. Most importantly the records contain a comparison of the actual sentence that the offender received with the mean simulated sentence. The standardized score of the actual sentence is used. The difference between the actual sentence and the mean simulated sentence divided by the standard deviation of the simulated sentences or standardized score (z-score) is used as an "index of disparity".

A SAS file is created from the OS file. The SAS file is then used to produce special reports describing potentially disparate cases.

Cases that are identified as potentially disparate by the ASR model are further scrutinized by the Board staff. If these staff believe a sentence to be disparately high or low, it is presented to Board Members for an official decision. A finding of disparity by the Board results in a letter of advice to the court.

Sentences cannot legally be lengthened, however several persons have experienced recall and subsequent shortening of their sentence as a result of the disparate review process. Twenty-four man years of reduction in sentence as of January 11, 1982 have resulted in state savings of at least $312,000.

It is anticipated that ultimately at least 20 persons per month will be found to have a disparate sentence.