

Proceedings of the Third Workshop on Law and Justice Statistics 1985

Edited by George G. Woodworth

Papers presented in Las Vegas, Nevada, August 3-4, 1985, at a workshop developed by the American Statistical Association Committee on Law and Justice Statistics.

Workshop sponsored by the Bureau of Justice Statistics and the American Statistical Association's Continuing Education Department with special acknowledgment to the Criminal Justice Statistics Association.



NCJRS

AUG 10 1988

ACQUISITIONS

112230

U.S. Department of Justice
National Institute of Justice

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

Permission to reproduce this copyrighted material has been granted by

Public Domain/BJIS
U.S. Department of Justice
to the National Criminal Justice Reference Service (NCJRS).

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

**U.S. Department of Justice
Bureau of Justice Statistics**

Steven R. Schlesinger
Director

This project was supported by Grant No. 82-BJ-CX-KO44, awarded to the American Statistical Association by the Bureau of Justice Statistics, U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. The project was monitored by Sue A. Lindgren of the Bureau of Justice Statistics. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official positions or policies of the U.S. Department of Justice or the American Statistical Association.

The papers and discussion in this proceedings volume are reproduced essentially as received from the authors. The papers in this volume have been reviewed by the editor but have not been subjected to a formal refereeing process.

NCJ-112230

The Assistant Attorney General, Office of Justice Programs, coordinates the activities of the following program offices and bureaus: the Bureau of Justice Statistics, National Institute of Justice, Bureau of Justice Assistance, Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime.

**American Statistical Association
Committee on Law and Justice
Statistics**

Robert M. Groves (1984-89),
Chair

Alan E. Gelfand (1980-86)
Chair (1985-86)

Charles H. Alexander, Jr.
(1987-89)

Lily E. Christ (1985-87)

Owen B. Dall (1987-89)

D. V. Gokhale (1981-86)

Wayne S. Fisher (1987-89)

Thomas A. Henderson (1983-88)

John P. Lehoczky (1986-88)

Colin K. Loftin (1984-89)

James P. Lepkowski (1986-88)

S. James Press (1980-85)

Albert J. Reiss, Jr. (1980-85)

John E. Rolph (1980-85)

G. Diane Saphire (1987-89)

Shanna H. Swan (1986-88)

Katherine Wallman (1984-86)

Herbert I. Weisberg (1986-88)

George G. Woodworth (1985-87)

Foreword

The Bureau of Justice Statistics (BJS) is pleased to have sponsored the third biannual "Workshop on Law and Justice Statistics" held in conjunction with the American Statistical Association's 1985 Annual Meeting in Las Vegas, Nevada. The workshop, organized and conducted by the ASA Committee on Law and Justice Statistics, provided a forum through which criminal justice statisticians and researchers, in both the private and public sectors and at all levels of government in the United States and Canada, could meet to exchange ideas and discuss current work. Improving such communication is fundamental to the BJS goal of improving the availability and dissemination of high-quality statistical information on crime and the criminal justice system.

By publishing these proceedings, we seek to share the papers with a much larger audience than the individuals who were able to attend the sessions. We believe that the potential benefits of peer interaction will be magnified many times as those who work daily with criminal justice statistics, whether in government, academia, or elsewhere, read the contributions of experts in various fields.

The Bureau's interest in supporting the work of the Committee on Law and Justice Statistics goes beyond biannual workshops. The committee represents a resource to the Bureau of extensive expertise in the field of criminal justice statistics, a resource that the Bureau frequently draws upon for a variety of purposes. Most significant, perhaps, is the access the committee gives us to methodological and statistical talents to address thorny issues in data collection and statistical application.

The relationship that has developed between the American Statistical Association's Committee on Law and Justice Statistics and the Bureau of Justice Statistics has proven to be a mutually satisfying one that allows the Federal Government to utilize the talents and experience of national experts in criminal justice statistics, allows comment from highly sophisticated users of BJS data, and provides input from those users and others into the design of data collection and analysis and into dissemination strategies to increase the utility of the data for a wide range of analytic needs. That relationship between the committee and BJS is an example of our commitment to produce high-quality data that meet the needs of all of their potential users.

Steven R. Schlesinger
Director
Bureau of Justice Statistics

The conference from which these proceedings emanate was conceived, designed, and arranged by the Committee on Law and Justice Statistics of the American Statistical Association. Our committee was created to provide an interface for the association with the legal, judicial, and criminal justice communities. As such, it attempts to:

- help disseminate information about law and justice statistics activities throughout the statistics community
- promote the development of quality statistical activities in civil and criminal justice settings
- consider and report upon relevant issues to guarantee the integrity of statistical programs maintained by the U.S. Department of Justice and other appropriate agencies and organizations.

Amongst our educational activities we have produced three biennial Workshops on Law and Justice Statistics, 1981 in Detroit, 1983 in Toronto, and 1985 in Las Vegas. These workshops were intended to bring together workers and researchers in this area who are interested in methodology. These workshops have generally been well received. This proceedings volume documents the most recent one. Our intent in publishing these technical papers is to extend the awareness of important methodological issues in law and justice statistics and to encourage greater interaction among the individuals from the diverse disciplines who are interested in this methodology.

We are currently witnessing a substantial growth in research at this statistical interface. The increase in activity level of our committee reflects this exciting circumstance. For 1987 we developed an intensive workshop on the National Crime Survey.

We continue to be grateful to the American Statistical Association and to the Bureau of Justice Statistics for their sponsorship and their strong emotional support.

Alan E. Gelfand
Chair, Committee on Law and
Justice Statistics of the
American Statistical Association

List of contributors

The authors, discussants, and workshop organizers are listed with their affiliations at the time the papers printed here were prepared. Their complete mailing addresses are presented in appendix B.

David C. Baldus
University of Iowa

John P. Lehoczky
Carnegie-Mellon University

Alfred Blumstein
Carnegie-Mellon University

Russell V. Lenth
University of Iowa

Lily E. Christ
City University of New York

Colin Loftin
University of Maryland

Piet de Jong
University of British
Columbia

Thomas J. Marx
Marx Social Science
Research, Inc.

Barry D. Gaudette
Royal Canadian Mounted
Police

Carla M. Noziglia
Las Vegas Metropolitan Police
Department

Alan E. Gelfand
University of Connecticut

Raymond Paternoster
University of Maryland

Samuel R. Gross
University of Michigan

Charles A. Pulaski, Jr.
Arizona State University

James J. Heckman
University of Chicago

John E. Rolph
The RAND Corporation

Thomas A. Henderson
Criminal Justice Statistics
Association

Steven R. Schlesinger
U.S. Bureau of Justice
Statistics

Joseph B. Kadane
Carnegie-Mellon University

George G. Woodworth
University of Iowa

Richard P. Kern
Virginia Statistical Analysis
Center

Paul F. Kolmetz
Virginia Statistical Analysis
Center

Preface

These proceedings are a record of presentations given at the third Workshop on Law and Justice Statistics held in conjunction with the Annual Meeting of the American Statistical Association (ASA) on August 3 and 4, 1985, in Las Vegas, Nevada. The goal of the three biennial workshops was to encourage statisticians, social scientists, and law and justice professionals to exchange information about problems in the law and justice system and about statistical research applicable to these problems; the fact that over 200 individuals have participated in these workshops suggests that they have achieved that goal.

Thanks are due to many people who helped to bring the third workshop together. They include other members of the Law and Justice Statistics Committee who organized workshop sessions: Lily Christ, Alan Gelfand, Thomas Henderson, Colin Loftin, and John Rolph. They include the discussants who kept the sessions lively: James Heckman and Joseph Kadane. Thanks go to Carla M. Noziglia of the Las Vegas Metropolitan Police Department who opened the forensic science session with a fascinating presentation of forensic science in action. Finally, many thanks to Sue Lindgren at BJS and Ede Denenberg and Jo Prezystas at ASA.

George G. Woodworth
Editor
Workshop Organizer

The papers and discussions in this proceedings volume are reproduced essentially as received from the authors. The papers in this volume have been reviewed by the editor but have not been subjected to a formal refereeing process. Statements or positions taken by the authors do not necessarily represent the views of either the Bureau of Justice Statistics or the American Statistical Association.

Contents

Foreword iii

List of contributors iv

Preface v

Thomas A. Henderson, Organizer
Steven R. Schlesinger

Richard P. Kern
Paul F. Kolmetz

Lily E. Christ, Organizer
Barry D. Gaudette
Piet de Jong
Russell V. Lenth

Alan E. Gelfand, Organizer

Alfred Blumstein
John P. Lehoczky

George G. Woodworth, Organizer
Colin Loftin, Organizer
James J. Heckman, Discussant
George G. Woodworth

Raymond Paternoster

James J. Heckman

Samuel R. Gross
George G. Woodworth
David C. Baldus
Charles A. Pulaski, Jr.

John E. Rolph, Organizer
Joseph B. Kadane, Discussant
Thomas J. Marx

Statistical tools for law and justice administration

Redesign of the National Crime Survey and the
Uniform Crime Reporting Program 1

Some observations on the development of objective
measures to aid decisionmaking in the administration
of justice 6

Statistics and probability in forensic science

Evaluating associative forensic science evidence 20

Bayes theorem in forensic science 26

On identification by probability 29

**Statistical studies of the law and justice system:
Criminal careers**

Quantification and modeling of criminal careers 41

**Statistical studies of the law and justice system:
Racial discrimination in capital sentencing**

Recent studies of race and victim effects
in capital sentencing 55

Racial discrimination and arbitrariness
in capital punishment: A review
of the evidence 59

Comments on three studies on disparity in capital
sentencing by race of victim: Baldus, Woodworth
and Pulaski; Gross and Mauro; and Paternoster 70

A reply to Professor Heckman 73

Reply to Professor Heckman 75

Jury selection

The mechanics of random juror selection 78

Appendix A: Program 83

Appendix B: Addresses of speakers and organizers 85

Redesign of the National Crime Survey and the Uniform Crime Reporting Program

Steven R. Schlesinger

U.S. Bureau of Justice Statistics

Abstract

The Bureau of Justice Statistics (BJS) has sponsored projects to redesign the two major programs that collect data on crime and crime victims in the United States: BJS' National Crime Survey (NCS) and the FBI's Uniform Crime Reporting (UCR) Program. The NCS redesign project, begun in 1979, was a total reassessment of the design, administration, and potential uses of the survey. It was undertaken by a consortium of experts in criminology, survey design, and statistics, with the active participation of BJS and the Census Bureau, which serves as the collection agent for the NCS. Major features of the NCS design, administrative procedures, and analysis conventions were examined, and a large body of material was prepared as a basis for recommendations on sample design, collection procedures, questionnaire content, comparability with the Uniform Crime Reports series, utilization, and analytic and processing needs.

The UCR Program redesign began in 1982 with a study of the existing program by a private contractor. It was overseen by a joint BJS/FBI Task Force and was guided by a steering committee made up of police practitioners, researchers, academicians, the media, and representatives of the leading law enforcement professional organizations. The study examined the original program as begun in 1930 based on the plan of the Committee on Uniform Crime Records of the International Association of Chiefs of Police, the current program as operated by the FBI, and alternative potential enhancements to the current UCR system. A set of recommendations was developed and published in Blueprint for the Future of the Uniform Crime Reporting Program in June 1985.

This paper describes the history, development, recommendations, and implementation of these two redesign efforts. It has been updated to discuss the current status of the two programs at the time of publication of these proceedings.¹

National Crime Survey redesign

In 1975 the National Academy of Sciences (NAS) was asked to evaluate the victimization survey program. Their comprehensive findings concerning ways in which the program should be redesigned are presented in Surveying Crime, 1976. Among the NAS recommendations were the following:

- More NCS resources should be devoted to "delineation of product objectives, to managerial coordination, to data analysis and dissemination, and to a continuing program of methodological research and evaluation."
- The NCS should produce "not only nationwide and regional data, but, on the same timetable, estimates for separately identifiable Standard Metropolitan Statistical Areas (SMSA's) and for at least the five largest central cities within them."

- The NCS screener, that is, that part of the questionnaire that ascertains whether the respondent has been a crime victim, needs to be drastically altered to increase its effectiveness in prodding respondents' memories and to minimize its complexity.
- Additional questions need to be added to allow measurement of independent variables important for understanding the dynamics of crime victimization. These would include questions dealing with ecological factors, victim characteristics, lifestyle, and protective or preventive measures.
- "A major methodological effort on optimum field and survey design for the NCS should be undertaken."

Following the academy's evaluation, an internal review of the NCS program was begun. In conjunction with this review, a conference was held in 1978 to discuss topics and priorities for a 5 year research program on national victimization survey statistics, and a study of the utility and benefits of the National Crime Survey was conducted. These assessments indicated a need for intensive examination and subsequent redesign of the NCS. In 1979 a contract was awarded to the Bureau of Social Science Research (BSSR) to begin this work. BSSR headed a consortium of experts in criminology, survey design, and statistics who contributed to various phases of the project. In addition, the project received guidance from an advisory panel drawn from the criminal justice, statistics, and social science communities.

- The NCS redesign project was charged with investigating a wide range of issues, including--
- improving the accuracy of recall for victimization incidents
 - expanding the scope of crimes covered
 - increasing cost effectiveness
 - enhancing the analytic utility of NCS data
 - improving utilization of NCS data.

Implementing changes of this scope to the NCS is necessarily a complex task. BJS has adopted a two-stage phase-in strategy for changes to minimize disruption of the series. The first phase, implemented in July 1986, introduced improvements that are expected to have minimal impact on measured crime rates and offer immediate benefits in the utility of NCS data. The second stage is scheduled for phase-in beginning in January, 1989, and ending in July, 1991, and will introduce changes that are likely to have an effect on crime rates. This second phase of the redesign recommendations will result in a "break" in the series. BJS is exploring methods to minimize and document the magnitude of the break, primarily by developing a statistical "splice" between old and new data. However, in many cases comparisons of data collected before and after the phase-in will remain difficult. These changes, nonetheless, will result in more efficient collection of NCS data, greater accuracy of victimization estimates, and improved opportunities for analysis of victimization-related issues.

Questionnaire revisions. The NCS instrument is divided into two sections. A "screener" collects background data on all respondents and includes a number of questions designed to elicit reports of criminal victimization experienced by the respondent or the household that has occurred during the 6-month reference period prior to the interview. The other section of the questionnaire is an "incident form," which gathers data on the characteristics of each reported incident and its consequences. A number of changes are planned for both parts of the questionnaire. Many of these were adopted during the first phase of implementation, although most alterations to the screener will occur during the second phase because of their potential for affecting victimization rates. Questionnaire changes introduced in 1986 included--

- expansion of questions on the victim's use of self-protective measures
- the addition of a question about drug use by the offenders
- the addition of questions about the victim's contacts and experiences with the criminal justice system.

Improving the accuracy of recall for victimization incidents. Any survey such as the NCS that relies on respondents recalling experiences they have had during a recent timeframe can only be as accurate as respondents' memories. The NCS redesign has placed a major emphasis on discovering ways to prod respondents' memories so that more accurate reporting of victimization events can be achieved. Four large field tests of various "screening" strategies were undertaken during the life of the project, and the techniques that proved most fruitful will be incorporated into the second phase of changes to the survey. These changes to the NCS screener are expected to produce an increase in measured crime rates because of improved measurement techniques. The new screening strategies will allow BJS to publish more accurate crime measures, and they may also allow a sample reduction because of the expected increase in the number of victimizations reported. This would minimize data collection costs while still collecting enough crime victimization data to produce reliable estimates.

Expanding the scope of crimes covered. Since its inception, the NCS has collected data on rape, personal robbery, assault, personal and household larceny, burglary, and motor vehicle theft. The NCS redesign consortium devoted part of its efforts to an investigation of ways in which the scope of crimes measured by the NCS could be expanded. Among the possibilities investigated were bombings, parental kidnaping, arson, fraud, and vandalism. A number of the crimes studied did not appear to be promising for measurement using victim survey methods because of the rarity of the crime or concerns about the potential unreliability of victim reports. Vandalism appeared to be the most promising addition to the survey, but several measurement difficulties had to be overcome before it could be included as a regular NCS crime type. One problem is that personal and household vandalism must be distinguished from other types of vandalism, such as damage to common areas in apartment buildings or damage to objects in one's neighborhood, such as street signs. Another difficulty is that the NCS crime incident form is not appropriate in many ways for the measurement of this crime, and alternate ways to collect incident data had to be developed. In addition, many vandalism incidents, such as damage to screens or windows, may be confused with

attempted burglaries, which could result in a reduction in reports of attempted burglaries, solely as a result of such a questionnaire change.

Deletion of current NCS items. Taken together, the questionnaire changes described above would result in a substantially longer NCS interview, with negative consequences for data collection costs and respondent burden. To compensate, a number of items traditionally in the questionnaire either were eliminated in Phase I or will be removed in Phase II. Some questions, such as the long battery dealing with unemployment and attempts to find work, are to be deleted permanently. Other items may be included only in periodic supplements. Among these are questions dealing with medical and property insurance coverage, recovery and/or repair of stolen or damaged property, and time lost from work as a result of an incident. Decisions on questionnaire cuts are guided by a desire to maximize the analytic utility of the data collected at every interview, to maintain useful time series, and to collect enough data on rare events to make reliable analysis possible. Questions to be deferred for supplemental administration were judged either to be relatively stable over time, so that detection of trends would not be compromised, or to involve frequent enough responses so that periodic administration would still provide adequate cases for analysis.

Longitudinal design. The most ambitious innovation under discussion for the revised NCS is adoption of a true longitudinal design during the second phase of implementation. The NCS has had some features of a longitudinal survey since its inception, in that rotating panels of households are interviewed for seven successive interviews at 6-month intervals. However, no attempt has been made to retain in sample those respondents who move, and attempts to link NCS records have been performed post hoc by independent researchers for special purposes. Retaining in sample those respondents who move and introducing a longitudinal processing system that facilitates the linkage of records will allow use of more powerful statistics for calculating annual change estimates and will also enhance the long-term representativeness of a population-based NCS sample, thereby reducing error in these estimates. In addition, introduction of such a design will allow analysis of important issues for the first time including--

- whether crime victimization is a factor in the geographic mobility of respondents
- the long-term health and economic consequences of victimization
- victims' contacts with the criminal justice system over an extended period of time
- victims who experience one-time, periodic, or relatively continuous victimizations and also the factors--such as the type of crime and victim or offender characteristics--that vary across these different temporal patterns
- the degree to which respondents victimized in one year also account for victimizations in other years.

A final decision has not been reached on implementation of a longitudinal capability for the NCS, largely because cost estimates and feasibility studies are not yet complete. However, introduction of this change will provide important benefits for the analytic utility of the NCS, if it proves fiscally realistic.

Improving utilization of NCS data. One of the major goals of the redesign was to enhance the value of NCS data to a wide range of users. In addition to efforts to improve the content of the NCS questionnaire and to facilitate new types of analysis, the redesign has taken a number of steps to broaden the scope of applications for the series. This work was facilitated by input from the project advisory panel and from a panel of practitioners assembled to provide advice on these concerns.

A major criticism of the NCS program has been its failure to provide data for specific States and localities. Because of the stratified probability sample employed, the survey collects data in only a limited number of locations. However, the major problem in releasing State and local data has been the Title XIII restrictions under which the Census Bureau operates. This statute is designed to protect the confidentiality of data collected by the decennial census, and any sampling frame, such as that for the NCS, that relies on decennial data is also covered by these restrictions. Problems arise in the release of subnational data because combining this information with demographic information about respondents on the public use files could result in the identification of particular respondents. The Census Bureau has released NCS tables annually for the largest States and has performed special analyses for subnational areas on request, but this procedure is costly and time consuming and does not allow the user direct access to the data to investigate different issues that may arise as he or she becomes better acquainted with the data.

BJS plans to address some of these difficulties by releasing NCS files aggregated at the State and county level and for major cities, beginning with the 1987 data year. Sampling data on the public use files will be scrambled to prevent a match with particular respondents, thereby dealing with confidentiality concerns. These files will contain key NCS variables and important economic and demographic data for the appropriate geographic unit. We also hope to include Uniform Crime Reports (UCR) data for corresponding jurisdictions. Release of such files will allow BJS to deal swiftly with requests for data on particular subnational units and will allow users some analytic flexibility in investigating victimization patterns for the areas of interest.

To facilitate use of these files, BJS is currently investigating their release in a form compatible with microcomputers. In addition to the data files, the release would include menu-driven software dedicated to analysis of the data with routines for "generic area" modeling. Many subnational units are not represented in NCS files, and this capability would allow users to estimate victimization levels for areas not represented in NCS data by using information from areas with similar characteristics. The utilization panel assembled for the redesign project was particularly helpful in developing a typology for this purpose, which includes 14 subnational area types and an additional residual category. Classification was based on population, land use (urban, rural nonfarm, and rural farm), MSA status, and incorporation of the geographic unit. It is not yet clear when the aggregated files could be released in this form, but we hope that it can follow shortly after the Census Bureau releases the initial files.

Telephone interviewing. The NCS Redesign Consortium evaluated previous research conducted by the Census Bureau on the effect of telephone interviewing and other relevant research related to the impact of telephone interviews and recommended that the amount of telephone interviewing be increased.² The plans under consideration involve conducting the first interview at a household in person and then all but one subsequent interviews by telephone, to the extent possible, (that is, if the respondent has access to a phone, is willing to accept a telephone interview, and can be reached by phone for the scheduled interview).

Computer Assisted Telephone Interviewing (CATI). CATI technology played a major role in testing revisions to the NCS, in that all tests conducted by the Survey Research Center at the University of Michigan utilized CATI vehicles. This technology involves programming the questionnaire into a computer and flashing screens containing questionnaire items onto a monitor for interviewers to read during the interview. Responses are entered at the interviewer's keyboard during the interview and become part of the record for that interview. This procedure offers a number of advantages in the collection and processing of questionnaire data:

- o Interviewers work out of a centralized facility, with supervisors present. Supervisors can unobtrusively monitor interviews in progress and detect problems in interview practices.
- o CATI software can be programmed to reject obviously erroneous codes, thereby reducing the likelihood of interviewer error in keying data.
- o Because skip patterns are programmed, the possibility of interviewers skipping over required questions is greatly reduced. This CATI feature makes possible the development of more complex instruments than would be possible with a paper questionnaire.
- o Because no hard copy is involved, the keying of paper instruments to a computer record is eliminated in data processing, thereby saving time, reducing costs, and eliminating a step in which processing errors may be introduced.
- o Because CATI interviewing need not be conducted with interviewing personnel who reside in the same area as the respondent, CATI facilities may be located at sites where the wages available for interviewers are attractive. This makes possible the recruitment of a higher quality interviewing staff, reduces turnover, and thereby minimizes training costs.

While CATI allows some economies over face-to-face interviewing in its reduction of field and some data processing costs, it does require additional expenditures in other areas. A site for an interviewing facility must be acquired and developed, computing hardware must be requisitioned, and software must be written. In addition, some interviews scheduled for CATI administration may not be completed due to failure to reach respondents over the telephone. Such cases must be recycled back to regional offices for administration in person by regular field interviewers. Finally, two processing systems must be developed, one for CATI interviews and one for non-CATI interviews. Data collected by both collection modes must ultimately be merged to create a single data set.

While this technology has potential for reducing NCS costs, definitive data on the impact of CATI data collection do not yet exist. After several years of development work, the Census Bureau began rigorous testing of CATI on actual NCS cases in January 1987 to

determine its impact on data quality and cost. These tests are also designed to determine the feasibility of recycling uncompleted CATI cases back to regional offices for personal interviews. Data on CATI performance are currently being studied, and additional analyses will be performed as more data are collected. A final decision on implementation will be made in time to begin CATI interviewing for all eligible respondents who are administered the new NCS questionnaire in January, 1990.

Increased use of supplements. Another major innovation planned for the NCS is to begin administering supplements on a more regular basis. These will include scheduled supplements to collect data at regular intervals that are not deemed essential for regular inclusion in the NCS. The remainder will be one-time supplements on crime-related topics for which the NCS would be an appropriate vehicle. Thus, in addition to providing higher quality data for estimation of crime levels and analysis of various attributes of criminal victimization, the NCS will soon be able to serve as an omnibus survey for crime-related topics, thereby enhancing its utility as a major data source for criminal justice policymaking research.

Uniform Crime Reporting Program redesign

The International Association of Chiefs of Police (IACP), aware that the Uniform Crime Reporting (UCR) Program had remained essentially unchanged since its beginning in the 1930's, passed resolutions requesting that the FBI, as the operator of the national program, conduct a review of the UCR Program. A joint FBI/BJS Task Force was established to oversee the review, and Abt Associates of Cambridge, Massachusetts, was selected as the contractor to perform the work. In addition to the joint FBI/BJS Task Force, the contractor was guided by a steering committee consisting of representatives of State UCR programs, local law enforcement agencies, law enforcement professional associations, the media, researchers, and other users of UCR data.

The redesign project consists of three phases:

- o Phase I documented the history and evolution of the program and developed an exhaustive set of issues identified by police, researchers, planners, and the media.
- o Phase II examined the issues identified in Phase I, studied alternative potential enhancements to the UCR, and concluded with a set of recommended modifications.
- o Phase III, the current and final phase, covers the selection and implementation of the recommended changes.

Recommended modifications. The recommendations developed in Phase II are contained in a report entitled Blueprint for the Future of the Uniform Crime Reporting Program. The report was released in June 1985 with an invitation for public comment. By September 1985 approximately 100 letters had been received, with the overwhelming majority of them containing an endorsement of the study's findings.

- The major recommendations of the report are to--
- o convert the UCR system to a two-level reporting system under which most agencies report basic offense and arrest information similar to that currently reported (Level I), while a comparatively small sample of agencies report much more extensive information (Level II)
 - o convert the entire UCR offense reporting system to unit-record reporting in which local law enforcement agencies submit reports on the characteristics of each individual criminal incident (for example, location, time, and presence of weapon)
 - o convert the entire UCR arrest reporting system to unit-record reporting in which local law enforcement agencies submit reports on the characteristics of each individual arrest
 - o distinguish attempted from completed offenses
 - o distinguish among crimes against businesses, crimes against individuals or households, and crimes against other entities
 - o institute routine, ongoing audits of samples of participating UCR agencies in order to establish the extent of error in the system on a continuing basis for both Level I and Level II
 - o develop the UCR, NCS, and Offender-Based Transaction Statistics (OBTS) systems as independent programs providing complementary criminal justice statistics for multiple purposes; the strengths of each of these data systems should be continued and enhanced rather than compromised to achieve superficial comparability
 - o continue efforts to provide the means for reconciling UCR and NCS data by evaluating seriousness scoring and by preparing periodic publications, special studies, and technical documentation
 - o support continued and improved user services, including a user data base with files linked over time, the ability to draw samples of offenses for analysis either by the UCR staff or by outside researchers, and improved response to public queries.

Implementation of the redesigned UCR. Testing of definitions and procedures began in 1986, and implementation began in fiscal year 1987 on a phased basis. Specific data element definitions, coding instructions, and incident reporting form revisions were developed, and an award was made to the State of South Carolina to test the revised definitions, instructions, and form revisions and their utility for capturing the expanded data elements. South Carolina is also developing computer software for data entry and tabulation. Data collection commenced in nine South Carolina local police departments in April of 1987. South Carolina is regarded as a "demonstration State"; it was selected to determine whether it is possible to collect the data that have been proposed as constituting a redesigned UCR.

The implementation guidelines and automated data capture specifications that were provided to South Carolina were revised in July 1987 to reflect changes that became necessary as that pilot project progressed. Additional changes arising from a review and actual usage of the July specifications were incorporated in the guidelines in February 1988. Copies of the guidelines are sent to all State and local UCR Programs to keep them apprised of the progress of the redesign effort and to encourage their continued involvement in the redesign process. Final guidelines and data processing specifications are expected to be issued on or about July 1, 1988.

At about the same time as the South Carolina test was beginning, the BJS Director wrote a letter to each Governor, describing the effort to make the first major change in UCR in more than 50 years and indicating the availability of Federal support. A full program announcement describing the availability of fiscal 1987 grant funds was sent to each State UCR Program. Eighteen States applied for funding, and 13 awards were made. A second set of seven applicants was selected for funding in fiscal 1988. The awards range from \$18,000 to \$390,000; the only prohibition on the funds is that they cannot be used for computer hardware procurement, consistent with the legislative history of the Justice Assistance Act of 1984.

In April of 1987 a meeting was held in St. Louis with the 13 States that received UCR awards as well as representatives of various national associations interested in the UCR redesign. The participants were taken step-by-step through the reporting guidelines and data processing specifications and given the opportunity to ask questions and raise issues. Similar meetings will be held in the future so that the States can exchange information with each other and with BJS and the FBI about problems encountered and solutions devised.

Anticipated benefits of a redesigned UCR. The new crime reporting program will vastly increase the amount and quality of information available to local police administrators, policymakers at all levels of government, the research and academic community, and the public. Specifically, it will--

- provide more detailed information for developing strategies for crime control
- provide data on victim, as well as offender, characteristics
- afford law enforcement agencies the opportunity for crime pattern analysis, with the inherent benefit of improved patrol force allocation
- provide data on crimes for which data traditionally have been lacking, namely drug-related offenses, sex crimes, family violence, and child abuse
- permit rapid analysis of particular crime and criminal justice issues of special concern to the police, policymakers, and the public
- provide the data necessary to better reconcile the findings of the UCR with the NCS
- provide a means of quickly spotting errors and reporting inconsistencies through routine quality assurance edit checks built into the new software for both the national and local programs
- make it possible to collect, tabulate, and correlate much more information about criminal events without significantly increasing costs associated with use of hard-copy media (paper reports) currently in use by most departments. This will be accomplished through computer technology linked with interactive input devices.

References

Groves, Robert M., Peter V. Miller, and Velma J. Handlin. "Telephone Survey Methodology: A Review." Manuscript, 1982, Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.

Federal Bureau of Investigation (1987). "Redesigned Uniform Crime Reporting Program Draft Implementation Guidelines and Automated Data Capture Specifications." (photocopied)

Penick, Bettye K. Eidsen (editor) (1976). Surveying Crime. National Academy of Sciences, Washington, D.C.

Poggio et al. (1985). Blueprint for the Future of the Uniform Crime Reporting Program: Final Report of the UCR Study.

Taylor, Bruce (forthcoming). Redesign of the National Crime Survey (working title). Bureau of Justice Statistics, Washington, D.C., USGPO.

Notes

¹ Portions of this material have been presented elsewhere by BJS staff. For more detailed information on the NCS redesign, consult Taylor, Bruce, Redesign of the National Crime Survey (working title). For more detailed information on the recommendations for the UCR, consult Poggio et al., Blueprint for the Future of the Uniform Crime Reporting Program: Final Report of the UCR Study.

² For a review of the relevant telephone interviewing literature and discussion of the methodological issues facing the NCS program in implementing telephone interviewing, see Groves, Robert M., Peter V. Miller, and Velma J. Handlin, "Telephone Survey Methodology: A Review" Manuscript, 1982, Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.

Some observations on the development of objective measures to aid decisionmaking in the administration of justice

Richard P. Kern and Paul F. Kolmetz
Virginia Statistical Analysis Center

Introduction

Many objective decision-making tools used in the criminal justice field are a product of a statistical analysis of historical practice whose findings are tempered in varying degrees by normative input from those policy officials charged with designing the tool. Two examples of such instruments are objective risk assessment scales and objective sentencing guidelines.

Objective risk assessment instruments which attempt to predict the future likelihood of offenders failing along some dimension (e.g., failure to appear in court, new criminal behavior) are, in theory, largely predicated on the assumption that there are patterns in these failures that can be uncovered by a close scrutinization of their characteristics. Thus, by studying historical cases and their outcomes, one could presumably construct a probability table of sorts (risk assessment matrix) which predicts the likelihood of failure for a given offender on the dimension of interest. Objective sentencing guidelines usually do not attempt to predict risk but rather are largely an explicit portrait of historical practice that is instituted to ensure more equity and consistency in all future sentence dispositions.

The process of creating these so-called objective decision-making tools, then, relies heavily upon our ability to statistically identify "relevant" variables from the wide myriad of factors commonly made available to criminal justice decision-makers. Doing this in a reliable and valid fashion is quite often an arduous task due to the serious limitations of much existing criminal justice system data. It is therefore important that researchers working in this area communicate their ideas and experiences so that others who are contemplating the development of these tools fully realize the dimensions of the task. In this vein, this paper will relate some of our experiences in developing objective decision-making tools for use at diverse stages of the criminal justice system.

Objective Pretrial Risk Assessment Instruments

Objective Pretrial Risk Assessment Instruments are designed to inform a judge or magistrate of the relative likelihood of defendants failing to appear (FTA) for scheduled court dates or becoming involved in pretrial crime if they are released back into the community. The great majority of jurisdictions which have adopted objective risk assessment tools for use at pretrial screening however, have done no more than slightly modify or rotely institute a risk scale created back in the 60's by the VERA Institute of Justice for use in predicting pretrial failures

in The Manhattan Bail Project. This widespread uncritical and verbatim adoption of a scale whose generalizability has not been demonstrated (Bohnstedt and Geiser, 1979; Kirby, 1977; Clarke, 1983) is often the result of "real world" constraints rather than a belief in the external validity of the tool adopted from another study. Developing an objective pretrial risk assessment tool from scratch consumes a great deal of time and resources--both rare commodities in most criminal justice agencies.

In Virginia, a statewide Risk Assessment Task Force made up of judges, magistrates, sheriffs and corrections officials concluded that their fellow practitioners would be unreceptive to utilizing a risk instrument not premised on data drawn from Virginia offenders. However, mindful of the resources required for a statewide study, the task force chose to first undertake a pilot study of the concept in one Virginia locality.

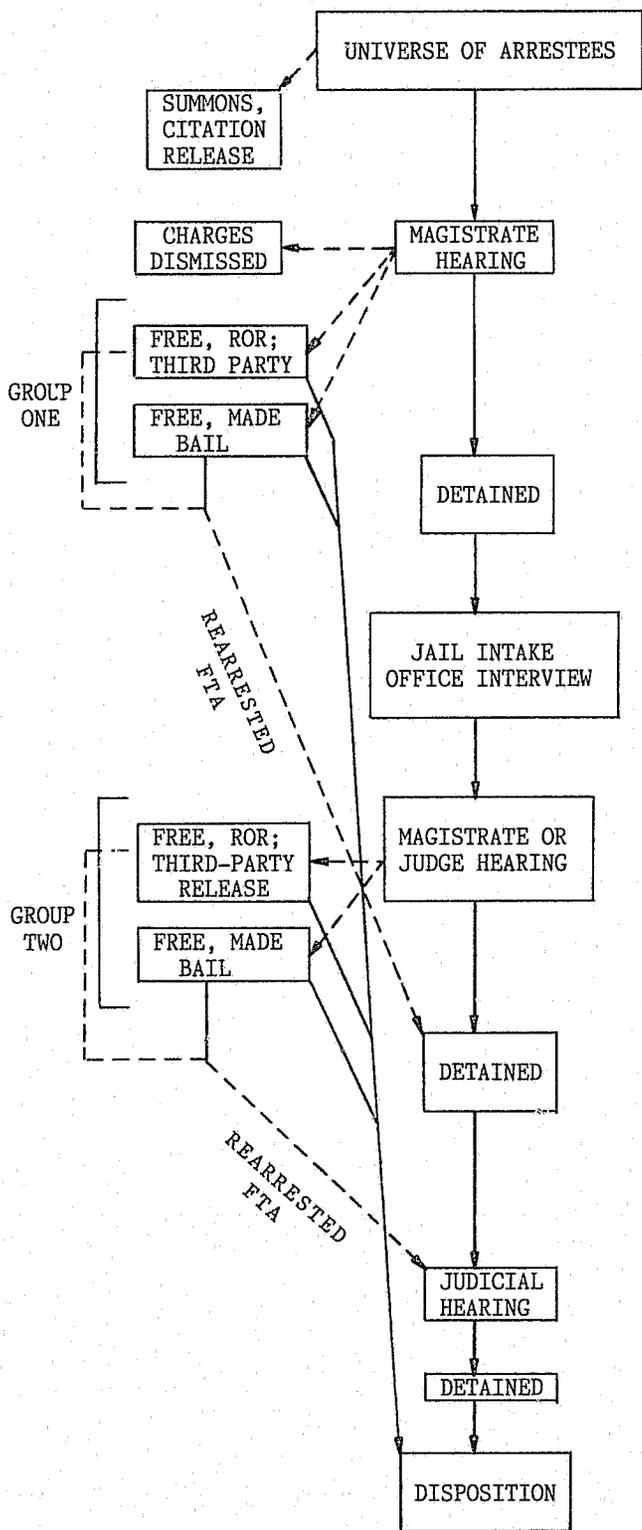
The site selection process for this study focused on two principal criteria: (1) would the local criminal justice decision-makers agree to use an objective risk instrument and, (2) did the locality have the historical data necessary to support the research work. While the first criterion limited our pool of sites somewhat, the second one almost eliminated every site in the state. Very few sites collect detailed systematic information on all arrested defendants. In Virginia, if a defendant is not ultimately convicted of a felony charge, chances are there will be only minimal paperwork containing very sketchy information on the offender and the case.

Ultimately, a Northern Virginia city was chosen which did have local officials willing to cooperate and which also had a jail intake office that collected a great deal of information on a portion of all arrested defendants. Unfortunately, this locality only interviewed defendants if they were unable to obtain their release after a hearing before a magistrate and subsequently had to spend at least one night in the local jail. Though an oversimplification, Figure 1 visually demonstrates the process involved after a defendant is arrested in this Virginia city and the problems posed by relying only on the information collected by the jail intake office.

Ideally, in conducting this research one would draw a probability sample from the universe of all defendants released pretrial and not just a subset thereof. The complete exclusion of defendants released prior to a jail interview would introduce systematic bias into the study. The reality of the situation, though, is that any cases sampled from Group 1 would be ultimately discarded from any statistical analysis due to excessive missing data on important research

FIGURE ONE

PRETRIAL CASEFLOW FOR ARRESTED DEFENDANTS IN SAMPLED VIRGINIA LOCALITY



variables. Consequently, our study has drawn its sample from only the interviewed group (Group 2). Thus, the principles of methodological rigor have been compromised somewhat due to the limitations and constraints imposed by the available criminal justice system data.

What are the dangers posed by selecting this sampling strategy? If the patterns of pretrial misbehavior differ significantly between those defendants interviewed and those not interviewed, then any risk assessment tool devised from a study focusing only on the former group will be inadequate in predicting failures in the latter class of offenders. As a precautionary measure, our research has been expanded to include a general look at the characteristics of the pretrial failures in the non-interviewed group to ensure that no dramatic differences exist.

Another sampling problem encountered in conducting this type of risk assessment research concerns what has been termed the "low base rate problem" (Go'tfredson, 1974; Kirby, 1979). This refers to the statistical problems posed by having an insufficient number of cases which result in pretrial failures. In practice, the percentage of defendants who either FTA or become involved in pretrial crime rarely is more than 15% (Thomas, 1976; Kirby, 1979; Pryor, 1980; Lazar Institute, 1981). Thus, if one were to predict that all defendants released pretrial would succeed in avoiding pretrial misconduct, we would expect to be correct in at least 85 percent of the cases. Consequently, risk assessment tools must be highly discriminating to improve upon this figure. To compensate for this expected problem, our study oversampled the cases which involved either an FTA or pretrial criminality. In technical terms, the sampling routine employed was disproportionate stratified random sampling. Stratifying our sample by pretrial misconduct required us to first identify the universe of pretrial failures.

Identifying pretrial criminality proved to be a difficult and time-consuming process. We initially explored the possibility of drawing this information from prior record "rap sheets." This proved to be unfeasible for two reasons: (1) we would have had to request rap sheets on all interviewed defendants and sift through what amounted to an unmanageable number of cases and, (2) even if the rap sheet reflected a new arrest occurring while the defendant was on pretrial release, there would be no way of ascertaining if this was a result of new criminality or was rather the execution of an old outstanding warrant.

The strategy ultimately selected focused first on whether any defendants were reinterviewed in the jail within a 90-day period of an initial interview. A new interview would always be conducted if the defendant was rearrested for a new crime. Since some of the initial cases were disposed of within 90 days, not all reinterviewed defendant had, in fact, committed pretrial crime; however, this was a good starting point from which to identify possible "hits." Also, this method did not capture the occurrence of pretrial criminality in other jurisdictions. Fortunately, it was

manually search the entire court docket. Unfortunately, while it was easy to identify whether a case involved an FTA, it has turned out to be extremely difficult to ascertain whether the missed appearance was a deliberate act. There are many reasons, some legitimate, why defendants do not appear for scheduled court dates (e.g., illness, car trouble, got lost). Since all missed court appearances, regardless of origin, are a disruption to a court's operation, one might reason that, pragmatically, it is only necessary to know when they occur and not why. Were our study just interested in reporting the aggregate rate at which defendants FTA, this logic would suffice. However, we are interested in predicting FTAs. Intuitively, it seems likely that most technical FTAs are random in nature and, therefore, are difficult, if not impossible, to predict. Deliberate FTAs, though, do not have accidental causes and, perhaps, are characterized by enough similarities to allow the development of a statistical probability model that predicts this phenomenon.

Theoretically, then, it makes sense to separate out technical FTAs from those which are deliberate. Despite our study's access to all court, jail, and police files, we have met with very limited success in measuring the reasons for an FTA. Kirby (1979) has suggested that deliberate FTA be defined as an instance where the defendant does not voluntarily appear for a scheduled court appearance a month after a bench warrant or capias has been issued. While such a definition can potentially be applied to cases where the reasons for an FTA are missing, it does require knowing whether the reappearance was voluntary. This information, also, has often eluded our data collectors.

The difficulty in dissecting the character of FTAs will probably adversely affect the predictive efficiency of the risk assessment tool to be constructed. One strongly suspects that the unimpressive predictive ability of pretrial risk scales (explained variance rates (R^2) ranging from a high of only .16 to a low of .02, Kirby, 1977) has been principally due to this inability to properly operationalize the dependent variable. In spite of the poor statistical efficiency of these models, most of the risk assessment tools devised from them have still been somewhat successful in increasing pretrial release rates without any concomitant increase in FTAs or pretrial crime (Beaudin et al., 1981; Pryor, 1980; Toborg, 1981; Clarke, 1983).

Most of the difficulties we've experienced in this risk assessment project derive from the poor quality of criminal justice system data. The problem is that we often must rely on data collected and maintained by criminal justice agencies which are not designed to support research--particularly risk assessment research. This situation confronted Virginia authorities when the decision was made to initiate the research necessary to implement a probation risk assessment instrument.

Probation Risk Assessment Research

A probation risk assessment instrument is designed to predict the likelihood of offenders failing if placed on probation. The objective behind the adoption of such a tool would be to place as many offenders as possible on probation, thus alleviating jail and prison crowding, without experiencing any concomitant increase in probation failures (revocations). In Virginia, both judges and probation officers were seen as being the primary users of such a tool. Theoretically, those who pose significant probation risks would receive from a judge either an incarcerative sentence or a probation term accompanied by numerous strict conditions. Probation officers would use the instrument to prioritize their large caseloads so that the greater risks receive closer supervision.

The first inclination of those charged with developing this instrument was to review existing state-maintained data bases to determine if any could support this type of study. A great deal of information was being gathered on some offenders but little of it was relevant to this type of research. The temptation, however, was still great to somehow adapt this easily available information to fit our needs. Due to limitations in both time and resources, we must oftentimes conduct our research with data not ideally suited towards producing the desired end product. Given the immense dimensions of the projected use of the probation risk assessment instrument, it was decided that every effort must be made to ensure that it be reliable and valid. Accordingly, the policy choice was made to initiate a new automated information system which could supply the data elements necessary to support statewide risk assessment research.

Since there was already in existence a source document prepared on all probationers which contained a good deal of the information desired, it was targeted for revision to make it completely amenable to both risk assessment research and automation. This document was the Pre-Sentence Investigation Report (PSI).

PSIs contain very detailed information regarding an offender's prior record, demographic characteristics, instant offense, social activities and lifestyle, and court case circumstances. PSIs are compiled by a probation and parole officer usually at the request of a sentencing judge. The primary purpose of the PSI is to provide accurate and objective information to the judges so that they can arrive at the sentence which best fits the need of the offender and community. While the PSI is initiated at a judge's request, the information it contains is used by many others in the criminal justice system. For example, the PSI is used in Virginia by the Division of Adult Services in classifying and developing a treatment program for offenders; by the Parole Board in considering inmates for parole; by probation and parole officers in counseling and rehabilitation efforts during

supervision; and as a source of pertinent information for systematic research and development of new rehabilitative procedures.

The PSI is, then, a valuable source of information to many in the criminal justice system. PSIs, however, are typically done in narrative format which usually prohibits a reader of the document from immediately finding vital pieces of information. Additionally, in Virginia, as in many states, their format was loosely structured with no rigid guidelines, resulting in uneven treatment of offender and offense characteristics. It was decided that while the PSI was the optimum source of information for risk assessment research, its format would have to be altered to ensure a more consistent and complete treatment of each case.

A task force composed of probation managers, corrections officials, statisticians, information system specialists, and criminologists was formed to develop a standardized PSI format which would supply the information necessary to support the probation risk assessment research and, at the same time, still continue to meet the daily needs of its many users.

Within the task force there was a continuum of opinion about the extent of standardization that should be imposed on the PSI. At one extreme was the viewpoint that standard subject headings over narrative discussion would suffice. In contrast was the position that all information items be objectively devised into discrete check-off categories with no narrative section being included. The task force ultimately opted for a blend of these two positions. While it was essential that vital "risk" and "sentencing" items be objective, discrete, and easily identifiable, it was also understood that narrative space must be allowed to either cover items not included in the standard form, or amplify on those which are, but require further explanation. While a blend, the format decided upon would place stronger emphasis on boxed discrete items.

There are pros and cons to this style of PSI standardization. As noted beforehand, there are many users of the PSI, and necessary information is easier to locate when in the same position on the report (Maloney and Raymond, 1977). Standardization of information into objective and discrete categories of interest also enables decision-makers to appraise offenders in a more expedient and consistent fashion.

A standardized PSI of the type recommended by the task force also serves to highlight key information. Important information is often diluted by surrounding trivia in narrative PSIs (Keve, 1961). The highlighting of relevant information presented in a standardized and succinct fashion may also result in more consistent sentencing practices. Hogarth (1971), for instance, has suggested that one possible reason for sentencing disparities is the great variance in quality and quantity of information being provided to the judiciary.

An obvious advantage to redesigning PSI information elements into standardized categorical variables is that it allows for the easy automation of the data. A standardized PSI form can be so designed that data entry personnel can work directly from the form itself without the need for a secondary coding form.

Turning to disadvantages, one of the most often heard complaints regarding the standardized PSI format suggested by the task force was that it was not as "readable" as the narrative reports. Many surveyed probation officers felt that a PSI should be open-ended, attractive to the eye, and easy to read. The proposed format, they believed, would not be that.

Another oft mentioned criticism of the proposed format was that it encouraged misleading information by not presenting the "complete picture." The following selected comments from field officers aptly portray this concern.

This report format is so rigidly structured that it has the frightening potential to inhibit a probation and parole officer's evaluation of an individual offender as a person.

Eventually I can see that officers will become lazy and only put in requested data and not take the extra step that produces a quality report. In the future there will be no writing skills necessary for this job.

All in all, it is very difficult to fit the diverse characteristics of the human organism into neat little boxes.

The overall structure of the standardized format may result in a report consisting essentially of disconnected data that will not give the court or other users of the report insight into the dynamics/factors that resulted in the subject becoming an offender.

Along similar lines was the criticism that the probation officer's role was being significantly reduced to that of a mere data collector. Many probation officers have degrees in disciplines such as criminology, sociology, psychology, and social work, and feel adequately equipped to conduct in-depth analyses of the offenders they prepare reports on. Checking off boxes on a standardized PSI, they argue, strips them of this dimension of their job.

Standardizing the PSI also entails additional costs, most of which would not be incurred with the narrative-type report. First, there is the cost of printing new forms with a detailed instruction manual complete with appropriate appendices. Officer training sessions must then be organized and procedures established for training new recruits. Additional personnel will be required at the central office to screen the PSIs coming in from the field to ensure their completeness. Also, data entry work stations and personnel will be required. There may also be the expense of hiring programmers and analysts to

work with the established PSI data base if they are not already on staff. Finally, one must consider the development costs in establishing the automated PSI system. To date, Virginia has devoted well over 2,500 hours of staff time to the design of this system.

While it was felt that the pros of PSI standardization greatly outweighed the cons, there still was a tremendous amount of concern over the probation officers lack of receptivity to the new form. It was readily apparent that the input of the field officer was vital to the success of this effort; therefore, the task force was broadened to include someone to represent their perspective.

The actual process of moving from a narrative open-ended report to a more structured standardized one began by first conducting a content review of the information items contained in a "typical" narrative PSI. The depth of coverage in narrative PSIs was found to vary considerably, depending upon the type of offender and judge the report was being prepared for. It was considered important at this stage, though, to identify those items which were usually found in most PSIs regardless of the case circumstances. The next step was to identify any information items which were statutorily required on the PSI.

The task force then turned its attention toward identifying potential risk assessment items which should be included on the PSI. Since there had not been any risk assessment research conducted in Virginia, it was considered fruitful to review the models developed in other states which had done so. The committee members studied quite closely the much acclaimed risk assessment system that has been developed in Iowa (Fischer & Stageberg, 1980). While there is no evidence that the factors found to be significant in assessing offender risk in Iowa are generalizable to other states, such items provide a useful starting point. The task force also reviewed the risk assessment system being used in Wisconsin with the same purpose in mind (Wisconsin Case Classification Study, no date).

A literature review of the criminological literature as it concerns risk assessment was also undertaken to ensure that no important data elements were being overlooked (see, e.g., Gottfredson and Gottfredson, 1980; Landis, Mercer, and Wolff, 1969; Maltz, 1981; Sechrest, White, and Brown, 1979; Williams, 1979). The goal here was to identify factors which theoretically hold promise as key predictors regarding the likelihood of an offender succeeding if placed on probation.

Input for suggested items on the PSI was also solicited from all users of the report in the criminal justice system. For example, a prison classification officer may need a certain information item, such as I.Q. score, in order to properly do his/her job. Consequently, input was solicited from judges, probation officers, classification officers, parole officials, prison case workers, and other researchers as well.

Perhaps the most important information element in a risk assessment data base is the measure of the criterion variable which in this case is success or failure on probation. Obviously, this is not a factor that is known at the time the PSI is being completed; therefore, provisions have to be made for this vital item to be captured at some future date and tied to the automated PSI data base. In order to ensure a successful match/merge operation when this information becomes available, it is important to include enough unique offender identifiers on the PSI data base. In Virginia's case, the social security number, the central criminal records exchange number, and FBI number of the offender were judged adequate for this purpose. All three numbers were recommended in the event that any of the other two might be missing. In those cases where these unique identifiers were missing, less unique items such as name, race and date of birth could be used in this operation.

Having identified data items contained in a typical PSI, potential risk assessment variables, information items requested by users, and elements which would facilitate future match/merging operations, the task force turned its attention toward the decision of which items to highlight (i.e., "BOX") and subsequently automate. Because it could not be anticipated with certainty which items would prove to be useful in both risk assessment and sentencing research, it was decided to box and key all identified data elements. In sum, 162 unique information items were targeted for data entry. While 162 data elements seem overly excessive for risk assessment purposes, it should be pointed out that 71 of these items simply measure the offense charges at arraignment and sentencing.

The next task dealt with developing the discrete categories of interest for each of the identified data elements. For example, the item "EMPLOYMENT RECORD" was conceptualized into 5 discrete measurement categories: 1) Regular, few changes; 2) Regular, many changes; 3) Irregular; 4) odd jobs only; and 5) no work record. A detailed instruction manual was developed to ensure that probation officers completed each item in an effective and reliable fashion. For example, a "regular Work Record with many changes" was defined as full-time employment over 75 percent of the time during the 2 years prior to the instant arrest involving 3 or more different jobs all with different employers. Other items were not captured in this categorical fashion; rather they were measured in their natural interval state (e.g., number of previous felony convictions).

The physical layout of the standardized PSI form was the next concern addressed by the task force. Some members favored placing all boxed items at the rear of the report with the narrative portions preceding it. The principal objection to this suggestion was that the information on a particular matter would be scattered all over the report. The favored approach was the grouping of boxed items under general headings such as "Current Offense," "Criminal History,"

FIGURE TWO

SAMPLE PAGE FROM VIRGINIA'S STANDARDIZED
PRESENTENCE INVESTIGATION REPORT

MOST SERIOUS OFFENSE INFORMATION		MOST SERIOUS OFFENSE		OFFENSE CODE (VCC)
DATE OF OFFENSE (MM/DD/YY)	NO. OF CODEFENDANTS	RESISTING ARREST CHARGE NO <input type="checkbox"/> YES <input type="checkbox"/>	TYPE OF OFFENSE PERSON <input type="checkbox"/> 1 PROPERTY <input type="checkbox"/> 2 OTHER <input type="checkbox"/> 3	
LEGAL STATUS AT TIME OF OFFENSE ESCAPE <input type="checkbox"/> 0 INMATE <input type="checkbox"/> 1 MANDATORY PAROLE <input type="checkbox"/> 2 DISCRETIONARY PAROLE <input type="checkbox"/> 3 PROBATION <input type="checkbox"/> 4 BOND <input type="checkbox"/> 5 RELEASED SUMMONS <input type="checkbox"/> 6 RELEASED RECOGNIZANCE <input type="checkbox"/> 7 OTHER <input type="checkbox"/> 8 NONE <input type="checkbox"/> 9				
WEAPON USE NONE <input type="checkbox"/> 1 USED TO INJURE <input type="checkbox"/> 2 USED TO THREATEN <input type="checkbox"/> 3		WEAPON TYPE FIREARM <input type="checkbox"/> 1 KNIFE <input type="checkbox"/> 2 EXPLOSIVE <input type="checkbox"/> 3 SIMULATED WEAPON <input type="checkbox"/> 4 OTHER <input type="checkbox"/> 5 NA <input type="checkbox"/>		
OFFENDER'S ROLE IN OFFENSE ALONE <input type="checkbox"/> 1 LEADER <input type="checkbox"/> 2 ACCOMPLICE <input type="checkbox"/> 3 NOT DETERMINED <input type="checkbox"/> 4				CURRENT ARREST DATE (MM/DD/YY)
MOST SERIOUS OFFENSE VICTIM INFORMATION (CRIME AGAINST PERSON) NA <input type="checkbox"/>		INJURY TO VICTIM DEATH <input type="checkbox"/> 1 SERIOUS PHYSICAL <input type="checkbox"/> 2 PHYSICAL <input type="checkbox"/> 3 EMOTIONAL <input type="checkbox"/> 4 THREATENED <input type="checkbox"/> 5 NONE <input type="checkbox"/> 6		
VICTIM RELATIONSHIP TO OFFENDER NONE <input type="checkbox"/> 1 FRIEND <input type="checkbox"/> 2 FAMILY <input type="checkbox"/> 3 POLICE OFFICER <input type="checkbox"/> 4		PHYSICALLY HANDICAPPED VICTIM NO <input type="checkbox"/> 1 YES <input type="checkbox"/> 2 UNKNOWN <input type="checkbox"/>		VICTIM INFORMATION SEX _____ RACE _____ AGE _____
VICTIM IMPACT STATEMENT REQUESTED YES, ATTACH TO LAST PAGE OF PSI. NO <input type="checkbox"/> YES <input type="checkbox"/>				
NARRATIVE OF CURRENT OFFENSE				

"Marital/Residential Stability," with open space blended in for any narrative required within each designated heading. An example illustrating one page from the new PSI format in Virginia is found in Figure 2.

The new PSI form has now been in use in Virginia for approximately 6 months. At this juncture, two observations seem especially noteworthy regarding our experiences. First, probation officers are having a difficult time adjusting to the new objective format. Consequently, we've found that the narratives accompanying the boxed information items are often very lengthy and, by themselves, would constitute a rather complete report. Because they are in essence doing two reports, the amount of time it takes to prepare a PSI has understandably increased. Though many users, particularly judges, often complained about wading through pages of irrelevant narrative to locate the facts on a case, they still insist on this manner of coverage on their PSIs despite its redundancies. It is hoped that as time goes on judges will find that they seldom need to refer to these lengthy narrative discussions and can adjust the style of the reports they order.

The second observation about the new PSI form is that the manner in which prior record is measured has generated a considerable amount of feedback from the field. Many probation officers believe these items are too numerous, too time consuming, and of questionable value. There are, however, quite legitimate reasons for the extensive treatment devoted to the measurement of prior adult record.

Prior criminal record is, perhaps, the most crucial factor in the development of both objective risk assessment tools and sentencing guidelines. Though prior record is a seemingly inherent objective phenomenon, much subjectivity enters into its measurement. Most prior record measures depend upon "rap sheets" as their information source. Rap sheets are notorious for containing missing information, especially on dispositions and for both missing and double entries.

Perhaps the most troublesome aspect of rap sheets, though, is that they convey criminal charge information in a very general fashion oftentimes utilizing what are known as National Crime Information Center (NCIC) codes. NCIC codes are tied to very general offense descriptions (e.g., Rape--Gun) and were designed to be utilized across the states so as to ensure standardized crime information. While these codes enhance the capability to do interstate crime analysis, they do so to the detriment of intrastate offense research. Since many of the NCIC code descriptions map poorly to the Virginia statutes, most offense information in our state is communicated using the most general "catch-all" NCIC codes (i.e., Sex Assault). Because the Virginia statutes contain many unique variations of these general crimes, a great deal of information is lost when these codes are utilized. For example, the Code of Virginia contains 15 unique statutory variations of "rape" which vary across six degrees of statutory seriousness.

To remedy this problem, new crime codes were devised which mapped directly to the Virginia statutes. Because some prior record information is missing essential offense details, it was still necessary to provide some catch-all crime codes within this new scheme. However, since the specific codes are used to communicate all new offense convictions and those priors that do contain the required detail, the foundation has been laid for a more complete computerized picture of the nature of prior criminal records in Virginia.

The adult prior record items measured within the standardized PSI are: 1) Number of prior felony sentence events, 2) number of prior felony convictions for crimes against person, property, drug crimes, other, 3) number of prior felony convictions for instant offense at conviction, 4) number of previous felon commitments, 5) five most recent and serious prior criminal adult convictions, 6) number of prior probations completed and revoked, 7) number of prior paroles completed and revoked, 8) number of prior incarcerations received under 1 year & 1 year or more, 9) last previous arrest date or release from confinement, and 10) number of prior misdemeanor convictions. Though seemingly redundant, each of the items measures a different dimension regarding prior criminal activity. For instance, prior convictions for a crime the offender is currently convicted of again, may be indicative of crime specialization frequently found in career criminals. The listing of the 5 most serious and recent convictions utilizing the new crime codes allows for the development of more sophisticated prior record measures that consider the statutory seriousness and specific nature of each offense. The objective is to measure prior record from several different perspectives, both qualitatively and quantitatively, and to then empirically determine which measure best predicts offender risk.

A best predictor in a statistical sense, however, is not always the best measure in a practical sense. The research which developed the Florida Sentencing Guidelines reveals a good example of this phenomenon.

Objective Sentencing Guidelines

Sentencing guidelines are an objective decision-making tool which seeks to ensure equity in decision-making by structuring judicial discretion within the confines of weighted "relevant" factors selected by the judiciary and their advisors. In the final product, sentencing guidelines often reflect a large amount of normative input, yet such input is closely guided by an empirical analysis of prior judicial sanctioning practices. This analysis of past practice isolates the statistically significant factors which determine sentence severity and estimates their relative importance.

A survey of functioning Statewide Sentencing Guidelines Systems revealed that while each is unique in the factors included in the guidelines, there is unanimous agreement regarding the inclusion of prior criminal record and the seriousness of the instant offense(s). However, while there

TABLE I
PRIOR ADULT RECORD ITEMS FOUND WITHIN STATEWIDE SENTENCING GUIDELINES OF 8 SELECTED STATES

PRIOR ADULT RECORD MEASURES	FL	MD	MI	MN	PA	SC	WA	WI
Prior adult felony convictions				X		X		X
Prior adult serious violent felony convictions							X	
Prior adult violent felony convictions							X	X
Prior adult non-violent felony convictions							X	
Prior adult high severity felony convictions			X			X		
Prior adult low severity felony convictions			X					
Seriousness of prior adult convictions (by statutory degree)	X				X			
Seriousness of prior adult convictions (none, minor, moderate, major)		X						
Prior adult convictions for offenses similar to the instant offense at conviction	X	X	X				X	
Prior adult misdemeanor convictions			X	X		X		
Prior adult convictions for offense against a person		X						
Prior adult parole/probation violations		X						
Prior adult driving while intoxicated convictions			X		X			
Legal status at time of offense	X	X		X		X		X
Legal status at time of arrest						X		

is a tremendous amount of agreement that prior record should be a factor in objective sentence guidelines, there is a notable lack of consensus regarding the manner in which it is best operationalized. Table I illustrates the great diversity of ways in which the states have chosen to objectively measure this important factor.

The research effort behind Florida's Sentencing Guidelines went to special lengths to uncover the measure of prior record that "best" modeled judicial sentence behavior. Initially, Florida's felony court judges completed a survey in which they rated the relative seriousness of statutory offense descriptions. A magnitude estimation scaling technique was employed which asked the judges to numerically rate the severity of the crimes in relation to the standard ("MODULUS") crime "Burglary of an Unoccupied Structure," which was assigned a score of 100. The resulting perceptual measure of offense seriousness reflected a significant amount of rater consensus and revealed fine gradations in seriousness not reflected within the cruder statutory seriousness scheme (Kern and Bales, 1983).

Crime codes specific to the Florida statutes and similar to those developed in Virginia were used to convey all prior convictions and the instant offenses for 5,069 felony conviction cases. The use of these codes allowed for the creation of four alternative measures of both prior record and the instant offense which varied considerably in their sophistication.

The first and simplest of these schemes measured the number of unique offenses at conviction and the number of prior adult felony offenses for which the offender was convicted. These measures are simple frequency indicators and do not take into consideration the relative seriousness of the offenses. They also represent the most common manner in which these items are conceptualized in sentencing studies, although it has been popular to measure prior record in even simpler terms as a dichotomous variable (see e.g., Baab and Ferguson, 1967; Pope, 1975; Lizotte, 1978).

The next two measurement schemes relied upon the statutorily defined seriousness system devised by the state legislature which can be weighted in two different manners. The first statutory measure was a simple ordinal level indicator that ordered crimes from high to low based on their statutory severity (7 = capital felony, 6 = life felony, 5 = 1st degree felony, 4 = 2nd degree felony, 3 = 3rd degree felony, 2 = 1st degree misdemeanor, and 1 = 2nd degree misdemeanor). The other statutory measure was an "ordered metric" scale (Maranell, 1979) that weighs crime severity by the statutorily defined maximum incarcerative penalty in years (30 = 1st degree felony, 15 = 2nd degree felony, 5 = 3rd degree felony, 1 = 1st degree misdemeanor, .167 (2 months) = 2nd degree misdemeanor).

The last measurement scheme used, and certainly the most sophisticated in both its derivation and application, applied the perceptually-based seriousness weights to instant offenses and all prior convictions. For example, a murder in the commis-

sion of a felony received a score of 1310; robbery with a weapon-755.5; dealing in stolen goods--190.2, etc.

Each of the four offense weighting schemes used in the measurement of the instant and prior offenses were then tested independently, along with a host of other variables commonly used in sentencing research, for their impact on sentence decisions. In essence, then, four independently derived statistical models of sentence decision-making were generated each time a different crime conviction group used to stratify the sample (e.g., murder, sex crimes, robbery) was analyzed.

In general, the results of our analysis revealed that when instant and prior record offense variables were relatively important in the modeling of the sentence decision, the predictive efficiency of the statistical model could be improved dramatically by increasing the measurement sophistication of these key variables. For example, the explained variance (R^2) contribution of just these two offense variables in the analysis of violent personal crimes was .04 when the simple frequency measure was applied; .09 when the ordinal level statistical weights were used; .17 when the ordered-metric weights were assigned; and .21 when the perceptual weighting scheme was employed. This pattern, however, did not hold up consistently across all offense groups analyzed. In most cases, the offense weighting scheme based upon statutory maximum penalties (ordered metric) performed almost as well as the perceptually derived measure.

With the data analysis completed, the Sentencing Commission reviewed the various options for incorporating an offense seriousness scheme into the Guidelines. The perceptually-based scheme was favored by some because 1) it was normatively derived, 2) it made intra-statutory crime seriousness distinctions, and 3) it did often emerge as the best empirical predictor of past sentence decisions. However, these advantages had to be weighed against the drawbacks of using this measure. Most notably, the perceptual seriousness measure was more complex than the other alternative measures and would be awkward to use in a sentencing guidelines scheme. While the commission wished to be guided by the data, their major concern was that the final product be as elegant and parsimonious as possible. A sentencing guideline scheme that is unnecessarily complex and cumbersome to use and understand will likely be unacceptable to most judges.

The research staff explored several options for incorporating the perceptual seriousness weighting scheme into the guidelines in a manner which would not alienate potential users. One option involved the development of a computer software package which would internally calculate all guideline sentences based upon the relevant case characteristics. Under this scheme, a computer terminal and software package would largely replace the traditional guideline's worksheet that is manually completed by the state attorney handling the case. Aside from the considerable hardware costs involved, this proposal was rejected because many commission members viewed

TABLE II
FLORIDA MULTIJURISDICTIONAL SENTENCING
GUIDELINES FOR ROBBERY OFFENSES

	POINTS	SCORE
1. PRIMARY OFFENSE AT CONVICTION		
1ST DEGREE FELONY	150	
2ND DEGREE FELONY	75	
3RD DEGREE FELONY	25	---
2. SECOND OFFENSE AT CONVICTION		
1ST DEGREE FELONY	150	
2ND DEGREE FELONY	75	
3RD DEGREE FELONY	25	
1ST DEGREE MISDEMEANOR	5	
2ND DEGREE MISDEMEANOR	1	---
3. THIRD OFFENSE AT CONVICTION		
1ST DEGREE FELONY	150	
2ND DEGREE FELONY	75	
3RD DEGREE FELONY	25	
1ST DEGREE MISDEMEANOR	5	
2ND DEGREE MISDEMEANOR	1	---
4. NUMBER OF COUNTS OF PRIMARY OFFENSE		
ONE	0	
TWO	29	
THREE OR MORE	48	---
5. PRIOR ADULT CONVICTIONS		
EACH CAPITAL FELONY	100	
EACH LIFE FELONY	100	
EACH 1ST DEGREE FELONY	60	
EACH 2ND DEGREE FELONY	30	
EACH 3RD DEGREE FELONY	10	
EACH 1ST DEGREE MISDEMEANOR	2	
EVERY 5 2ND DEGREE MISDEMANOR	2	---
6. PRIOR JUVENILE CONVICTIONS		
EACH LIFE CONVICTION	100	
EACH 1ST DEGREE CONVICTION	60	
EACH 2ND DEGREE CONVICTION	30	
EACH 3RD DEGREE CONVICTION	10	---
7. TYPE OF WEAPON		
NONE	0	
WEAPON OTHER THAN FIREARM	12	
FIREARM	24	---
8. EXTENT OF VICTIM INJURY		
NO INJURY, NO CONTACT	0	
NO INJURY, CONTACT MADE	17	
INJURY, NO TREATMENT REQUIRED	34	
INJURY, MINOR TREATMENT REQ'D	51	
INJURY, HOSPITALIZATION REQ'D	68	
DEATH	85	---
9. LEGAL STATUS AT TIME OF OFFENSE		
FREE, NO RESTRICTIONS	0	
UNDER SOME FORM OF RESTRICTION	60	---
10. ROLE OF OFFENDER		
ACCESSORY	- 40	
ALONE OR EQUAL INVOLVEMENT	0	
LEADER	40	---
	TOTAL:	---

this as a "Black Box" approach which would only further discourage judges from using guidelines. The commission wanted a Guidelines Scoring System which offered no mystery in that anyone could accurately calculate a score and clearly see where each point originated. Since the seriousness weighting scheme based upon statutory maximum penalties was also a strong historical predictor and was a parsimonious measure already understood by the judiciary, this method was ultimately incorporated into Florida's first version of Sentencing Guidelines.

Table II illustrates an example of how these seriousness weights were implemented into the Florida Multijurisdictional Sentencing Guidelines. The actual points assigned to each offense factor (instant & prior offenses) were derived largely from the statistical analysis (unstandardized tobit coefficients) but the intervals between the score values assigned for each increment in statutory seriousness are consistent with those for their statutory maximum penalties.

TABLE II (Continued)
FLORIDA MULTIJURISDICTIONAL SENTENCING
GUIDELINES FOR ROBBERY OFFENSES

ROBBERY OFFENSES	
COMPOSITE SCORE	SENTENCE
0-85	Probation - 36 mos incarceration
86-100	4 years (3-5 years)
101-125	6 (5-7)
126-150	8 (7-9)
151-175	10 (9-11)
176-200	12 (11-13)
201-250	15 (13-17)
251-325	20 (17-22)
326-375	25 (22-27)
376-400	30 (27-30)
401+	Life

TABLE III

FLORIDA STATEWIDE SENTENCING GUIDELINES FOR ROBBERY OFFENSES

I. Primary offense at conviction

Degree	Number of Counts				Number of Counts Above 4
	1	2	3	4	
Life	102	122	133	148	x 15 =
1st punishable by life	82	98	107	119	x 12 =
1st	70	84	91	101	x 10 =
2nd	50	60	65	75	x 10 =
3rd	34	41	44	54	x 10 =

II. Additional offenses at conviction

Degree	Number of Counts				Number of Counts Above 4
	1	2	3	4	
Life	20	24	26	28	x 2 =
1st pbl	17	20	22	24	x 2 =
1st	14	17	18	19	x 1 =
2nd	10	12	13	14	x 1 =
3rd	7	8	9	10	x 1 =
MM	1	2	3	4	x 1 =

III. A. Prior record

Degree	Number of Prior Convictions				Number of Counts Above 4
	1	2	3	4	
Life	100	210	330	460	x 130 =
1st pbl	80	168	264	368	x 104 =
1st	60	126	198	276	x 78 =
2nd	30	63	99	138	x 39 =
3rd	10	21	33	46	x 13 =
MM	2	5	8	12	x 4 =

B. Prior convictions for Category 3 offenses

Number prior convictions ___ x 25 =

IV. Legal status at time of offense

No restrictions	0
Legal constraint	17

V. Victim injury (physical)

None	0
Slight	7
Moderate	14
Death or severe	21

TABLE III (Continued)

FLORIDA STATEWIDE SENTENCING GUIDELINES FOR ROBBERY OFFENSES

Points	Recommended Range
34-53	any nonstate prison sanction Community Control or
54-65	12-30 mos. incarceration 3 years incarceration
66-81	(30-3 1/2) 4 years
82-101	(3 1/2-4 1/2) 5
102-121	(4 1/2-5 1/2) 6
122-151	(5 1/2-7) 8
152-183	(7-9) 10
184-229	(9-12) 15
230-295	(12-17) 20
296-357	(17-22) 25
358-417	(22-27) 30
418-453	(27-40) Life
454+	

developed based upon data relevant to statewide sentencing practices. In the course of developing these guidelines, a new sentencing commission departed significantly from their predecessors on how to treat offense seriousness.

In reviewing the multijurisdictional guidelines, the commission members felt that the seriousness score increments based upon statutory maximum penalties were too large for weighting the instant crimes and wished to see a new alternative scheme. They also believed that additional offenses at conviction should not contribute as much weight as that assigned to the primary crime.

Instead of arbitrarily assigning offense seriousness scores a priori and then examining their predictive efficacy, it was decided to let the statistical analysis determine the underlying structure of the seriousness factor. Dummy variables were created based upon the statutory degree of the instant offenses (primary and additional). For example, if a primary offense was a first degree felony it received a score of 1, if not, it received a value of 0. This was done for all possible statutory degrees that could fall within a given guideline's offense category. The resulting offense seriousness score values derived from the statistical analysis were then largely adopted by the commission members (Table III). These seriousness scores do not reflect the dramatic incremental increases that resulted when the statutory maximum penalty weights were used. Also,

These multijurisdictional guidelines essentially paved the way for the eventual development of statewide sentencing guidelines in Florida. Since the multijurisdictional guidelines were premised only on historical sentencing patterns in four selected judicial districts, this new project required that a new set of guidelines be

unlike the multijurisdictional model where additional offenses were treated equally to the primary crime, the statewide model scores the seriousness of these crimes at about one-fifth the weight assigned to the primary crime.

Another significant change from the multijurisdictional model involved the treatment of counts. In the multijurisdictional guidelines the counts of only the most serious crime were considered and their weight was independent of its statutory seriousness. The statewide commission decided that all counts should be scored and that their weight should be tied to the seriousness of the charge. Several different methods of incrementing score values for each additional offense count were explored. Since offense seriousness is rarely an additive phenomenon, it was agreed that the score assigned to one count should not be doubled for two counts and tripled for three counts. The incremental seriousness increase would have to be something less than this but large enough to make an impact in the sentence. The commission ultimately adopted a scheme which increments the score for number of counts at a rate determined by an analysis of historical practice.

The statewide commission also believed that prior criminal record did not contribute as much weight to the sentence as it should have in the multijurisdictional guidelines. The statewide guidelines were to reflect a stronger policy statement on the punishment of recidivists. Thus, while it was felt here that the seriousness weights for prior crimes should be based upon the statutory maximum penalties (same as within multijurisdictional model), the prevalent feeling was that additional convictions for crimes with the same statutory degree should count at least twice as much for the second conviction, three times as much for the third conviction, etc. In the multijurisdictional guidelines, each successive prior conviction within the same seriousness category received the same weight as that assigned to the first conviction.

In practice, the final statewide model weights each additional prior within the same statutory class at a slightly higher ratio than that originally suggested (2nd conviction weight = 2.2 X weight of 1st; 3rd conviction weight = 3.3 X weight of 1st, etc.). Unlike the decisions made on the weighing of seriousness for the instant charges, staff was largely unable to guide the commission on the optimum weights for prior record due to excessive missing data. In general, the factors selected and weighting schemes adopted in the statewide guidelines reflected a stronger amount of normative input from the commission than that witnessed in the multijurisdictional effort.

The other significant change in the new statewide guidelines is that the sentences reflect "real time" (minus only gain/good time) and are not subject to early parole release. Discretionary parole was not available to any offender sentenced after the new guidelines went into effect. Since the data base used to develop statewide guidelines reflected sentences imposed under an operat-

ing parole system, the historical sentence ranges were no longer directly applicable. Consideration had to be made for the reduction of sentence lengths that was historically accounted for by parole releases. Ignorance of this factor would almost certainly result in significant increases in the prison population.

Accordingly, staff accessed available Department of Corrections' records to try and estimate what the historic time served was for given categories of offenses and offenders. While this data could not be disaggregated specifically enough to allow for a direct comparison to the guidelines data, it did provide some general information to guide the commission. Ultimately, the selected sentence ranges largely reflected the policy members' normative judgments on what were considered to be appropriate prison terms.

TABLE IV

HYPOTHETICAL OFFENDER
SENTENCED UNDER TWO
SENTENCING GUIDELINES SCHEMES

CASE CHARACTERISTICS	MULTI-JURISDICTIONAL SENTENCING GUIDELINES SCORE	STATEWIDE SENTENCING GUIDELINES SCORE
PRIMARY CONVICTION - ROBBERY 2ND DEGREE (1 COUNT)	75	50
SECOND CONVICTION - 3RD DEGREE FELONY (1 COUNT)	25	7
PRIOR RECORD - 2ND DEG FELONY (2) 1ST DEG MIS-DEMEANOR (3)	60 6	63 8
UNDER LEGAL RESTRICTION AT ARREST	60	17
TOTAL POINTS	226	145
RECOMMENDED SENTENCE	15 YRS. (13-17 YR. RANGE)	6 YRS. (5 1/2 - 7YR. RANGE)

Since the multijurisdictional guidelines were developed on the premise of the continuation of existing parole practices, its sentence ranges are considerably greater than those found in the statewide version. For example, a hypothetical offender with the characteristics portrayed in Table IV would receive a guideline recommended sentence of 15 years in the multijurisdictional model but only 6 years in the statewide system.

Conclusion

This paper has discussed three (3) research efforts which address the development of objective decision-making tools for use at varied stages of the criminal justice system.

The pretrial risk assessment study highlights some unique problems encountered in attempting to develop such instruments given the serious limitations of available criminal justice system data. The process of developing a standardized PSI designed to support probation risk assessment research illustrates that a move to a more objective document is oftentimes met with tremendous resistance from field personnel. Finally, the discussion centering on sentencing guidelines demonstrates the role that statistics can play as a heuristic device in guiding the hard policy choices involved in designing objective decision-making instruments.

Projects such as these represent a unique opportunity for criminologists and statisticians to work in concert with policy makers in an effort to enhance both the efficiency and equity of decision-making in the administration of justice. The implementation of objective decision-making instruments, however, is no automatic panacea for all the problems they are designed to address. Objective decision-making tools that are hastily created or premised upon inappropriate or unreliable data may perform very poorly in practice even though their criteria are explicitly articulated and "objective." It is hoped that the observations presented herein will help guide others who attempt to undertake the development of objective decision-making instruments in the future.

References

- Baab, George W. & William R. Furgeson (1967). "Texas Sentencing Practices: A Statistical Study," Texas Law Review, 45:3, 471-503.
- Beaudin, Bruce D., Donald E. Pryor, and D. Alan Henry (1981). "A Proposal for the Reform of Pretrial Release and Detention Practices in the United States," Pretrial Services Annual Journal, Vol. IV, 68-102.
- Bohnstedt, Marvin and Saul Geiser (1979). Pretrial Release Sourcebook, Vol. 1 in Classification Instruments for Criminal Justice Decisions, National Institute of Justice, U.S. Department of Justice.
- Clarke, Stevens H. (1983). "What Do We Know About Bail?", unpublished manuscript, Institute of Government, University of North Carolina, Chapel Hill.
- Fischer and Stageberg (1980). The Iowa Offender Risk Assessment Scoring System, Iowa Statistical Analysis Center publication.
- Goldkamp, John S., Michael R. Gottfredson, and Susan Mitchell-Herzfeld (1981). Bail Decision-making: A Study of Policy Guidelines, National Institute of Corrections, U.S. Department of Justice.
- Gottfredson, Michael R. (1974). "An Empirical Analysis of Pretrial Release Decisions," Journal of Criminal Justice, Vol. 2, 287-304.
- Gottfredson and Gottfredson, "Screening For Risk: A Comparison of Methods," Criminal Justice and Behavior, Vol. 7, No. 3, 315-330.
- Hogarth (1971). Sentencing As A Human Process.
- Kern, Richard P. and William D. Bales (1983). "Measuring Offense Seriousness: Testing A Perceptually Based Scale," in Measurement Issues in Criminal Justice, Gordon P. Waldo (editor), Sage Research Progress Series in Criminology, No. 32, 15-36, Sage Publications, Beverly Hills, California.
- Keve (1961). "The Message in Mr. Pujo's Dream," 25 Federal Probation 11.
- Kirby, Michael P. (1977). The Effectiveness of The Point Scale, Pretrial Services Resource Center, Washington, D.C.
- Kirby, Michael P. (1979). Failure to Appear: What Does It Mean? How Can It Be Measured?, Pretrial Services Resource Center, Washington, D.C.
- Landis, Mercer, Wolff, (1969). "Success and Failure of Adult Probationers in California," Journal of Research in Crime and Delinquency, Vol. 6, No. 1, 34-40.
- Lazar, Institute (1981). Pretrial Release: A National Evaluation of Practices and Outcomes: Summary and Policy Analysis, Lazar Institute, Washington, D.C.
- Lizotte, Alan J. (1978). "Extra-Legal Factors in Chicago's Criminal Courts: Testing The Conflict Model of Criminal Justice," Social Problems, 25:564-580.
- Maloney and Raymond (1977). "A Standardized Presentence Report: One State's Response," 41 Federal Probation 40.
- Maranell, Gary M. (editor) (1974). Scaling: A Sourcebook for Behavioral Scientists, Chicago: Aldine Publishing Company.
- Pope, C. E. (1975). Sentencing of California Felony Offenders, Washington, D.C., United States Government Printing Office.
- Pryor, Donald E. (1980). "Significant Research Findings Concerning Pretrial Release," paper presented to the National College for Criminal Defense.
- Roth, Jeffrey A. and Paul B. Wice (1980). Pretrial Release and Misconduct in the District of Columbia, Institute for Law and Social Research, D.C.

Sechrest, White, Brown (eds.), (1979). The Rehabilitation of Criminal Offenders: Problems and Prospects, Washington, D.C., National Academy of Sciences.

Thomas, Wayne H. Jr. (1976). Bail Reform in America, University of California Press, Berkeley, California.

Toborg, Mary A. (1981). Pretrial Release: A National Evaluation of Practices and Outcomes, A National Evaluation Program, Phase II Report, National Institute of Justice, U.S. Department of Justice.

Williams, (1979) The Scope and Prediction of Recidivism, Institute for Law and Social Research, Washington, D.C.

Wisconsin Department of Health and Social Services, the Division of Corrections. "Case Classification/Staff Deployment Project Report No. 14."

Evaluating associative forensic science evidence

Barry D. Gaudette

Royal Canadian Mounted Police

Since forensic scientists deal with complex scientific evidence, the significance of which the layman has little or no understanding, it is natural and legitimate to attempt to evaluate the significance of forensic science evidence through the use of statistics. It can also be dangerous. This does not, however, mean that presentation of statistics should be avoided. Many things in life which are dangerous if abused are of great value when used properly. In my opinion the use of statistics in evaluating associative physical evidence falls into this category. Indeed, the benefits and necessity of applying statistical reasoning to evidential value determinations have been well documented (1). The existence of certain statistical pitfalls does, however, mean that forensic scientists must be extremely careful to use statistics properly. In this presentation, I will discuss four steps and enumerate some factors that can lead to proper statistical evaluation of associative physical evidence. I will also attempt to relate the factors in a value equation.

The first step towards a proper statistical evaluation of associative forensic science evidence is developing a conceptual framework for the role of associative physical evidence, and its components, in the judicial process. The purposes of the courts are first, to determine guilt or innocence and second, to take appropriate action such as sentencing. The determination of guilt depends on many considerations; among these are proof that a crime was committed, intent, mental condition, rules of law and evidence, witness credibility, establishment of prerequisite associations, alternative explanations for associations, and alibis. Forensic scientists from different disciplines may have evidence relevant to any of these factors. Trace evidence, however, is usually related to just one factor--establishment of associations.

Associative evidence is defined as that evidence which attempts to establish associations between any combination of the following: accused, victim, crime scene or weapon. Associative evidence can take many forms. In addition to the physical evidence forensic scientists are concerned with, associative evidence can be provided through eyewitness testimony as well as statements of suspects and victims. Accordingly, associative physical evidence is usually only one component of associative evidence which, in turn, is but one of many factors used in determining guilt or innocence.

Let us now examine some of the components of associative physical evidence. The first com-

ponent, identification, involves classification of the questioned evidence material (e.g. blood--human blood--type A, PGM-1 etc., or fibre--wool--light blue wool, etc.). The second component is comparison of the questioned evidence material to a known sample. In the third component, interpretation, the forensic scientist evaluates the significance of the evidence.

Another way of looking at these three components is to consider the thought processes a forensic scientist follows. On the basis of the results of an examination, a forensic scientist draws a conclusion which he or she then interprets in giving an expert opinion. Identification yields results; comparison leads to the conclusion; and interpretation produces the expert opinion.

An understanding of the preceding conceptual framework can assist us in the second step towards proper statistical evaluation of associative physical evidence--defining the fundamental question. Let us examine three general types of questions that, at first glance, may appear equivalent:

- 1) What proportion of the suspect population would have characteristics the same as those possessed by the questioned physical evidence?
- 2) What is the probability of a coincidental match between the questioned physical evidence and the known sample?
- 3) What is the value of the evidence in establishing a particular association?

These three questions can be related to the components of associative physical evidence as follows. Question one evaluates the significance of an identification and may be answered by population studies and data bases which provide frequency of occurrence data. Question two evaluates the accuracy or specificity of comparisons. However, if we wish to evaluate the significance of associative physical evidence, we must ask questions of the third type. Accordingly, the fundamental question to consider is: "What is the value of the evidence in establishing a particular association?". I will now attempt to develop equations to represent this value.

In the absence of physical evidence, the likelihood of association can be given by the ratio $P(A)/P(N)$, where $P(A)$ is the probability of association and $P(N)$ is the probability of non-association. When the physical evidence, E , is considered, the ratio becomes $P(A|E)/P(N|E)$. V , the value of the evidence in establishing association, can then be

determined by the ratio $[P(A|E)/P(N|E)] / [P(A)/P(N)]$, which is a measure of the extent to which the likelihood of association has been changed by the evidence.

Probability theory tells us that this last expression is equal to $P(E|A)/P(E|N)$; i.e. V is the ratio of the probability of the evidence, given association, to the probability of the evidence given non-association. This likelihood ratio is now commonly recognized as providing the best measure of V . A further advantage of the likelihood ratio approach is that it is a multiplier. The importance of this can be seen when we recall that physical associative evidence is but one component of association which, in turn, is but one component of guilt. By means of a multiplier, the value of the associative forensic science evidence can be easily placed in the total evidence picture.

At this point let us note that there are two possible states of nature with regards to association. Either there was some form of association (A) or there was not (N). Ignoring inconclusive results, there are two possible outcomes of the forensic scientist's examination—either the evidence indicated association (E) or it did not (\bar{E}). If the state of nature is A and the forensic scientist gives an opinion indicating E, the forensic scientist is correct. Similarly, if the state of nature is N and the forensic scientist says \bar{E} , he or she is also correct. However, if the state of nature is A and the forensic scientist says \bar{E} , a type I error or incorrect exclusion has occurred. If the state of nature is N and the forensic scientist gives an opinion indicating E, a type II error or incorrect association has occurred (see Table 1).

TABLE 1
ERROR TYPES IN ASSOCIATIVE FORENSIC
SCIENCE EVIDENCE

	State of Nature	
	A	N
E	Correct ($P = 1 - \alpha$)	Type II error ($P = \beta$)
\bar{E}	Type I error ($P = \alpha$)	Correct ($P = 1 - \beta$)

A = association
N = non-association
 \bar{E} = evidence of association
 \bar{E} = no evidence of association

If we now note the analogy of $P(E|A)$ to $1 - \alpha$, the probability of an incorrect exclusion (type I error) and the analogy of $P(E|N)$ to β , the probability of an incorrect association (type II error), substituting into our value equation we get $V = (1 - \alpha) / \beta$.

There are many causes of type I and type II errors in forensic science. Let us look first at type II errors since they are the most serious. One cause of incorrect associations is examiner error. This may be due to inadequate training, use of improper methodology, malfunctioning equipment, outdated or improperly prepared reagents, low natural ability, carelessness or corruption. I will term the probability of type II errors as a result of examiner error β_E .

A second component of β is β_C which is the probability of incorrect association due to coincidental matches. Taking blood as an example, β_C represents the probability of an innocent person being wrongly associated with a victim because of the coincidental occurrence of two separate events: 1) presence of a blood stain on his clothing, and 2) the blood stain being of the same type as the victim. Since both these events must occur to cause a coincidental match, β_C is the product of two sub-components, β_F and β_O . β_F , the relative frequency subcomponent, is dependent on the relative frequency of blood, fibre, glass, etc. types in the suspect population. β_O represents the probability of type II errors due to coincidental occurrence of other prerequisite events, such as the following:

- 1) one particular hair type (out of approximately 9 types on the scalp) being the one found in evidence (2),
- 2) a person having a blood stain on his or her clothing,
- 3) a particular textile material shedding fibres,
- 4) an object being damaged (e.g. pieces, paint chips or buttons missing).

β_X , the third component of β , arises from other explanations for the possibility that in spite of the evidence, there was no association. One such explanation is secondary transfer; that is, the perpetrator of a crime might have transferred hairs or fibres picked up from previous associations. Assume, for example, that while individuals A and B are drinking together in a bar, A loses a hair which is deposited on B's shirt. If B then commits a crime, a secondary transfer of A's hair to the crime scene could occur and result in evidence indicating that A was at the crime scene, when in fact he had been nowhere near it. Other explanations for β_X include contamination and deliberate planting of evidence. β , then equals $\beta_E + \beta_C + \beta_X = \beta_E + (\beta_F \cdot \beta_O) + \beta_X$.

Turning now to type I errors, α , the probability of incorrect exclusion, also has an examiner error component (α_E). Type I examiner errors can arise from the same factors listed previously for type II errors as well as from a forensic scientist failing to find trace evidence that was actually on exhibits such as clothing, bedspreads, car seats, etc.

The second component of α is the probability of incorrect exclusion due to coincidence, or α_C . Some factors that contribute to α_C include:

- 1) questioned objects that are atypical,
- 2) questioned objects that are incomplete or too small,
- 3) known samples that are unrepresentative,
- 4) known samples that are incomplete or insufficient in size.

α_X , the third component of α is the probability of incorrect exclusion due to other factors such as deterioration, improper collection or handling of evidence and changes in known samples that occur between the time the crime was committed and the time the suspect is apprehended. α , then = $\alpha_E + \alpha_C + \alpha_X$.

Substituting the components of α into our value equation we get

$$V = \frac{1 - (\alpha_E + \alpha_C + \alpha_X)}{\beta_E + (\beta_F \beta_O) + \beta_X}$$

In addition to this general form, let us look at some other forms of the value equation. First we will define V as the average value of the evidence of a specific type in establishing association.

$$V = \frac{1 - (\alpha_E + \alpha_C + \alpha_X)}{\beta_E + (\beta_F \beta_O) + \beta_X}$$

where the variables α_i and β_i are average values.

Then, let us consider V_p , the value of evidence of a specific type in establishing association in a particular case. When the results of a casework comparison indicate there was some form of association, α a priori becomes 0. Also, to obtain the value of the evidence in establishing association in a particular case, we must include case-related factors. The equation for V_p therefore, becomes

$$V_p = \frac{1}{a_E + (b_{\beta_F} c_{\beta_O}) + d_{\beta_X}}$$

The factors a, b, c and d are variable dependent on the circumstances of each particular case. Then $V_p = yV$, where the variable y can have any value from zero to infinity.

These equations represent an attempt to capture the relationships of some of the factors influencing the value of evidence. They have potential applications in five aspects of forensic science: court testimony, research, management, training and report writing. The remainder of this paper will deal primarily with the first of these.

The basic principle to be observed in setting up ideal guidelines for court testimony is that the best testimony provides the maximum possible amount of information about the value of the evidence in establishing a particular association. In order to obtain such an output, the proper inputs are required. These include: as much information as possible about the case, the results of the analysis, background scientific knowledge, an understanding of the concepts involved in the value equations, and an awareness of all assumptions made and their effect on the accuracy and precision of the value results.

Let us now use these ideal guidelines as standards in assessing the merits of some present approaches to providing interpretations of associative evidence.

Some forensic scientists present only results and no conclusions. A variation on this approach involves giving "could have" conclusions with no additional information. Those using this approach give the same conclusion in all cases regardless of circumstances and provide no interpretation. They make no mention of how the evidence relates to the attempted establishment of association. It can readily be seen that this method falls far short of our ideal guidelines.

In the second approach, qualitative terms such as probable, highly probable, and very highly probable are used in court testimony. This approach allows for testimony variation according to case circumstances and provides more information than the first method. However, in deciding on a qualitative term to use, different examiners make different assumptions and use criteria which they generally do not mention in court. Furthermore, the inexactness inherent in qualitative phrases leads to the danger that different people will interpret them in different ways.

Forensic Serologists use an approach which, on the surface, appears to overcome the problem of exactness. This involves quoting relative frequency statistics for various blood typing factors. When we look at the value equations, however, it can be seen that this approach has some shortcomings. Frequency tells only part of the story. In providing frequency data without any other information, the following assumptions are made:

- 1) the probability of incorrect association due to examiner error is negligible,
- 2) β_O , the probability of incorrect association due to coincidental occurrence of other prerequisite events, is not important,
- 3) no explanations other than association exist for the evidence, and
- 4) the population for which frequency data is quoted is representative of the suspect population.

With blood evidence, some or all of these assumptions can be reasonable or self-canceling. However, in many cases, at least one of them will fail. For example, when a victim's

blood is found on the accused, β_0 , the probability of a person having blood stains on his clothing becomes important.

When assumptions are not spelled out and their reasonableness substantiated, a jury is left with the false impression that the numbers given are precise and exact. Accordingly, the third step towards proper evaluation of associative forensic science evidence is clearly stating and substantiating all assumptions. Just as it is dangerous to let lay people draw conclusions of similarity on the basis of such things as comparisons of electrophoresis patterns or infrared spectra, it is dangerous to let lay people draw conclusions as to the value of evidence from frequency data alone. Furthermore, it is not consistent with other parts of a forensic scientist's testimony. We do not simply present our raw data and leave the courts to draw their own conclusions as to similarity. Why, therefore, should we present only raw relative frequency data and let the courts draw their own conclusions as to the value of the evidence?

Many other forensic disciplines are trying to emulate Serology's approach by establishing computerized data bases which they hope will enable court presentation of frequency data. Whereas in Serology all the assumptions can sometimes be reasonable or self-cancelling, in other disciplines the nature of the exhibit material will generally cause at least one of the prerequisite assumptions to fail. Presentation of frequency data on its own will then lead to a distorted picture of the value of the evidence, along with a false sense of exactness.

If a forensic scientist is not extremely careful, statistics presented in relation to identification will become generalized in the mind of a juror as being directly applicable to association or even guilt (3). This demonstrates the importance of the fourth step leading to proper statistical evaluation of associative physical evidence--placing the statistical answers in the proper context.

As a result of the many factors involved in determining evidential value, it is unlikely that the forensic scientist will have statistical data that is sufficient of itself to answer question three and exactly specify V_p . Statistics relating to questions one or two can still be used, provided the forensic scientist is careful to point out their place and limitations. Another alternative is to use what I have termed the "touchpoint approach" (4,5). In this approach, a qualitative assessment of the net effect on a particular case of the various evidential value factors is used to modify quantitative average probability statistics which serve as a point of reference. This approach has the following steps:

1) Be aware of the factors affecting V_p , and their relationships as indicated by the value equations.

2) Through research, obtain an estimate of $\bar{\beta}$ and hence \bar{V} . Depending on the experimental design, this can be either done directly or through separate determinations of the components of $\bar{\beta}$, individually or in combination. The type of evidence and the circumstances involved will determine which $\bar{\beta}$ factors will have a large effect on \bar{V} and which will have a negligible effect. If, for example, $\bar{\beta}_x$ were 1/50, while all other $\bar{\beta}$ factors were each about 1/5000, the limiting factor in determining \bar{V} would be $\bar{\beta}_x$ and the other factors could be safely ignored.

3) Use background knowledge, combined with knowledge of the case, to assess the factors involved in V_p . Qualitatively determine y through consideration of the effects of a , b , c and d .

4) In presenting evidence, use the estimate of \bar{V} as a touchpoint.

5) Relate \bar{V} to V_p through qualitative phrases such as much higher, higher, about the same, lower and much lower.

This approach is based on the principle that whereas V_p is too variable to enable meaningful statistics to be obtained, \bar{V} will remain constant for a given evidence type. Furthermore, since it is an average, \bar{V} is much less population dependent.

Although the touchpoint approach is not without drawbacks and limitations, this analysis shows that it is generally the best available means of conveying the value of physical evidence in establishing a particular association. Accordingly, the purpose of computerized data bases should be re-thought. Rather than providing exact frequency data for direct court use, computerized data bases should be limited to providing data upon which to base qualitative statements, a much easier and more productive task. This analysis also demonstrates the need to have research projects aimed at determining \bar{V} or the various factors (β_x , β_E , β_F , β_0) for all types of associative evidence designed and carried out.

In summary, an understanding of the conceptual framework for the role of associative physical evidence, and its components, in the judicial process can lead to formulation of the fundamental question: "What is the value of the evidence in establishing a particular association?". If the forensic scientist then ensures that statistical answers are placed in the proper context and that all assumptions are spelled out and substantiated, a proper statistical evaluation of associative physical evidence will be obtained. Such evaluations can then help forensic scientists make many important decisions. It is hoped that the concepts presented here will be of assistance in decision making and stimulate the thinking of statisticians, members of the legal profession and forensic scientists.

References

1. Eggleston, R., "Evidence, Proof and Probability," (Second Edition), Weidenfeld and Nicholson Ltd., London, 1983.
2. Gaudette, B. D., "A Supplementary Discussion of Probabilities and Human Hair Comparisons," Journal of Forensic Sciences, Vol. 27, No. 2, April, 1982, pp. 279-289.
3. Tribe, L. H., "Trial by Mathematics: Precision and Ritual in the Legal Process," Harvard Law Review, Vol. 84, pp. 1329-1393, 1971.
4. Gaudette, B. D., "Some Further Thoughts on Probabilities and Human Hair Comparisons," Journal of Forensic Sciences, Vol. 23, No. 4, October, 1978, pp. 758-763.
5. Gaudette, B. D., "The Significance of Forensic Hair Comparison Evidence," presented at the American Academy of Forensic Sciences Annual Meeting in Anaheim, California, February 24, 1984.

GLOSSARY OF TERMS

Type I errors - incorrect exclusions

Type II errors - incorrect associations

P(A) - probability of association

P(N) - probability of non-association

P(A|E) - probability of association given physical evidence E

P(N|E) - probability of non-association given physical evidence E

P(E|A) - probability of the evidence E given that there was association

P(E|N) - probability of the evidence E given that there was non-association

α - probability of a type I error when attempting to establish an association

β - probability of a type II error when attempting to establish an association

V - value of evidence of a specific type in establishing an association

\bar{V} - average value of evidence of a specific type in establishing an association

V_p - value of evidence of a specific type in establishing an association in a particular case

β_E - probability of incorrect association due to examiner error

β_C - probability of incorrect association due to a coincidental match being made

β_F - that sub-component of β_C which is directly related to the frequency with which a given questioned sample of a particular evidence type is found in the suspect population

β_0 - the other sub-component of β_C . Represents probability of incorrect association due to coincidental occurrence of other prerequisite events

β_X - probability of incorrect association due to other factors such as secondary transfer, contamination and deliberate planting of evidence

α_E - probability of incorrect exclusion due to examiner error

α_C - probability of incorrect exclusion due to coincidence

α_X - probability of incorrect exclusion due to other factors

$\bar{\beta}_E, \bar{\beta}_F, \bar{\beta}_0, \bar{\beta}_X, \bar{\beta}_E, \bar{\beta}_C, \bar{\beta}_X$ - average probabilities: Insert the word 'average' in front of appropriate definition

a - the number of times β_E is higher or lower than average, due to the circumstances of a particular case

b - the number of times β_F is higher or lower than average, due to the circumstances of a particular case

c - the number of times β_0 is higher or lower than average, due to the circumstances of a particular case

d - the number of times β_X is higher or lower than average, due to the circumstances of a particular case

y - the number of times the evidence in a particular case is more (or less) valuable in establishing association than is the average evidence of that type

SUMMARY OF EQUATIONS

$$(1) \quad v = \frac{P(A|E) / P(N|E)}{P(A) / P(N)}$$

$$(2) \quad v = \frac{1 - \alpha}{\beta}$$

$$(3) \quad \beta = \beta_E + (\beta_F - \beta_O) + \beta_X$$

$$(4) \quad \alpha = \alpha_E + \alpha_C + \alpha_X$$

$$(5) \quad v = \frac{1 - (\alpha_E + \alpha_C + \alpha_X)}{\beta_E + (\beta_F - \beta_O) + \beta_X}$$

$$(6) \quad \bar{v} = \frac{1 - \bar{\alpha}_E + \bar{\alpha}_C + \bar{\alpha}_X}{\bar{\beta}_E + (\bar{\beta}_F - \bar{\beta}_O) + \bar{\beta}_X}$$

$$(7) \quad v_P = \frac{1}{a\beta_E + (b\bar{\beta}_F - c\bar{\beta}_O) + d\bar{\beta}_X}$$

$$(8) \quad v_P = y\bar{v}$$

Bayes theorem in forensic science

Piet de Jong

University of British Columbia

Summary

This paper argues that Bayes Theorem is not as important as is often claimed in the evaluation of forensic evidence. Instead, the notions of sensitivity and "false alarm rating" provide an easily understood and communicated framework for measuring the value of evidence. The notions clarify the forensically important issues concerning the value of evidence. Some of these issues and their pitfalls are outlined.

1. The relevance of Bayes Theorem to forensic science

Bayes Theorem is a complete and utter mystery to most people involved in the legal process. This seems a disastrous state of affairs, since statisticians often claim Bayes Theorem is indispensable to a proper evaluation of evidence.

The future also seems bleak. There is little progress in selling the idea that Bayes Theorem is important. Participants in the legal process who hear of the theorem usually dismiss it as irrelevant. Even Probability Theory is rarely treated as a serious domain for legal practitioners.

Bayes Theorem is a statement regarding mathematical probabilities. The statement must be distinguished from its practical implications. This paper argues that the practical implications can be arrived at without Bayes Theorem. The implications follow from commonsense arguments. The arguments are couched in terms of the "sensitivity" and "false alarm rating" of a procedure. These two notions are easily understood and communicated. Together, they imply an evaluation of evidence analogous to what is practically implied by Bayes Theorem, but without the mystery and confusion surrounding Bayes result. Attention is focused on concrete issues and diverted away from mathematical intricacies.

Downplaying the role of Bayes Theorem does not mean that the theorem is in some sense wrong. Nor is it argued that the theorem is irrelevant. What is argued is that most practically important implications of the theorem can be arrived at without the mathematical apparatus of Bayes Theorem. This clears the way for Forensic scientists to move on and tackle the forensically important issues surrounding the evaluation of evidence.

2. Sensitivity and false alarms

"Sensitivity" and "false alarm rating" are crucial notions in the evaluation of forensic evidence. The following two examples explain these notions better than a long and involved discourse. Although simplified, the examples have the same essential features as more complicated and often messy forensic situations.

1. Fire alarms: A device built to serve as a fire alarm is found to signal 'fire' whenever there is fire. At first sight this appears an excellent device. Actually the device is worthless if it also often signals fire when there is none. For the device to be effective it must have two properties. First, it must, with high frequency, signal fire when there is fire. Second, it must rarely give false alarms.

2. Medical diagnosis: A medical test is designed to check for the presence or absence of a particular disease. In medical terminology, an effective test is one that is both highly sensitive and specific. Sensitivity refers to the frequency with which the test correctly signals the presence of the disease. Specificity relates to the frequency with which the test incorrectly signals the presence of the disease i.e. the frequency of false alarms.

The examples emphasize that the result of a particular test for a condition must be evaluated relative to two properties of the employed procedure. First, the propensity of the procedure to signal the condition when in fact it is present. This is called the sensitivity of the procedure. Second, the propensity of the procedure to give false alarms. For lack of another name, this is called the "false alarm rating" or simply FAR.

Sensitivity and FAR are properties of a procedure relative to an assertion (e.g. there is fire, the disease is present). A particular result or outcome of the procedure is judged in the light of the procedure's sensitivity and FAR. This discussion indicates five basic ingredients in the evaluation of a particular result.

1. A procedure used to generate a result
2. An assertion
3. A sensitivity rating for the procedure relative to the assertion
4. A FAR for the procedure relative to the assertion
5. A result generated by the procedure

These ingredients are contained in the following layout

likelihood of obtaining the evidence actually obtained with the procedure if assertion is

true	sensitivity
false	FAR

Suppose a fire alarm has a sensitivity of 99%. This means that if there is fire then the probability of the alarm going off is 0.99. This emphatically does not mean that if the alarm goes off then the probability of fire is 0.99. This kind of mistake is called the "fallacy of the transposed conditional". The trap is easy to fall into, and crucial to keep out of.

3. Sensitivity and FAR for forensic evidence

Forensic evidence is usually used to argue the identity of two objects or persons. For example, the evidence is that two persons have similar attributes such as blood, fingerprints, or hair. At issue is how much support is provided by the evidence.

Evidence of identical blood types, fingerprints or hair serve as a proverbial "alarm". The alarm purports to signal the identity of two persons. Alarms are only taken seriously if they are sensitive to the condition they purport to signal, and if there are few false alarms.

1. Blood tests: Evidence is given (the alarm is sounded) that an accused has the same blood type as that left at the scene of a crime. The evidence purports to signal the condition that the crime blood is the accused's blood. The degree of support for the assertion depends on the sensitivity and FAR of the blood typing procedure. The sensitivity of the procedure is the probability of type identity between the suspect and crime blood if the crime blood is from the accused. Barring blood typing errors, this probability is 1. The FAR of the blood typing procedure is the probability of blood type identity if the accused is not the source. The FAR will vary according to the blood type, and population of alternative sources.

2. Fingerprints: Fingerprints are widely used to implicate accused persons. As various authors have remarked, this is because in all the experience with fingerprinting, no two persons have been found to have the same print. Hence empirical experience indicates that the FAR of fingerprint identification is extremely low. On the other hand, the sensitivity of fingerprint identification is high: given adequate imprints from the same source, the probability of identification is high.

3. Hair evidence: Gaudette and Keeping (1974) estimated that with standard forensic techniques of hair identification, the probability of identifying two random hairs from two random different individuals is approximately 1 in 40,000. In other words, the standard procedure of forensic hair identification has a low FAR. In subsequent articles, Gaudette (1979) argued that hair identification procedures are sensitive: two random hairs from the same individual are, with high frequency, judged as coming from the same individual.

4. Oral testimony: Mr. Smith says that he saw Mr. Jones at the scene of the crime. The assertion to be tested is that Jones was at the scene of the crime. The sensitivity of the procedure is how likely Smith would say that he saw Jones at the scene of the crime if Jones was at the scene of the crime. The FAR of the procedure is how likely Smith would make this statement if Jones was not there. Both the sensitivity and FAR will be estimated from such things as the usual veracity of Smith, the circumstances of the oral testimony, the motives for Smith saying one thing or the other etc.

4. The weight or value of evidence

Single measures of the weight or value of a piece of evidence include

1. sensitivity - FAR
2. sensitivity / FAR
3. $\log(\text{sensitivity}) - \log(\text{FAR})$

The second measure is sometimes termed the 'likelihood ratio' or 'value equation' and is formally equivalent to the multiplier in Bayes Theorem.

All three measures emphasize that the weight of evidence depends on a comparison of sensitivity and FAR. All are reasonable indicators. It does not matter exactly which measure is used. Each has advantages and disadvantages. For example, the second measure rates equally the situations sensitivity=1, FAR=0.01 and sensitivity=0.00001, FAR=0.000001. Perhaps more reasonably, the first measure assigns more weight to the first situation.

The above measures are used to judge the value of evidence. They indicate the degree to which we should modify our beliefs as to the truth or falsity of an assertion. The measures do not establish absolute standards for "large" or "small" values. Nor do they assign absolute probabilities to events such as 'identity' or 'guilt'. Absolute probabilities of this sort are difficult to determine in all but trite circumstances. Even if determined, they provide only a limited basis for action. This is well illustrated with the example where n spectators watch a show. It is known that only one has paid--all others gained illegal entry. At the end, one of the specta-

tors is randomly chosen and charged. The accused is guilty with probability $1-1/n$. However, this arbitrarily high probability can never provide the basis for conviction.

It is instructive to analyze the example from the viewpoint of the weight of evidence. The evidence is that only one in n persons paid the entrance fee. To determine the sensitivity and FAR of the evidence we have to conceptualize a set of possible evidential outcomes of which the obtained evidence is but one. Little thought is required to convince oneself that these calculations are meaningless. Accordingly, the evidence has no direct implication to the assertion of 'guilt' or 'innocence'.

5. Calculating sensitivity and FAR

Sensitivity and FAR are relative frequencies over a number of real or conceptual experiments. The actual numbers are inevitably estimated. Quantitative people prefer a reasonably rigorous determination of these estimates. Non quantitative types tend to deal with sensitivity and FAR in a vague and intuitive manner.

The source of the estimates may be experience, sampling studies, or enumerations of whole populations. The estimates may be wholly or partly supported by mathematical or scientific arguments.

However derived, the estimates may be imprecise, incorrectly determined, endowed with too much precision, irrelevant to the situation at hand, or misinterpreted. All these dangers are real. However, such shortcomings can never justify the abandonment of the concepts of sensitivity and FAR.

6. Pitfalls in the use of sensitivity and FAR

The use of sensitivity and FAR is subject to many pitfalls. A common mistake is to consider only sensitivity or only FAR, not both. As the fire alarm example illustrates, this is meaningless. Measures of sensitivity and FAR are only meaningful when compared to each other. Of course in certain situations it is common knowledge that one or other is near one or zero. However it is crucial to keep both in mind.

A second mistake is to keep the relevant assertion vague and in the end identify it with "guilt". Guilt usually involves a whole host of notions beyond mere association or identity. The assertion with respect to which sensitivity and FAR are estimated are specific and are usually only loosely connected to the notion of guilt.

The "fallacy of the transposed conditional" is a usual third mistake. This is to for example identify sensitivity with the probability of truth the assertion rather than the probability of obtaining the evidence given the truth of the assertion. Often this fallacy is com-

bined with the previous mistake in which case for example sensitivity is interpreted as the "probability of guilt".

A fourth mistake relates to the fact that both sensitivity and FAR are relative frequencies with respect to a number of real or conceptual experiments. These experiments or cases must mirror the case at hand. For example, suppose a fire alarm is rated at a sensitivity of 95% and FAR of 1%. These readings correspond to the performance of the alarm over a number of situations--for example standard house fire situations. The readings may be totally irrelevant to the performance of the alarm in a chemical factory environment. Accordingly, sensitivity and FAR figures are only relevant to a particular case to the extent that the case is analogous to the cases providing the estimation basis for the quoted figures.

Hair evidence is an example of how sensitivity and FAR figures are sometimes misused. The 1 in 40,000 figure quoted above corresponds to the conceptual experiment of drawing two hairs at random from two different random individuals. The figure indicates that, on average, hair evidence has a low FAR relative to the assertion of identity. However, the figure cannot be uncritically used as the appropriate FAR reading in a particular situation. First, in a crime situation, the crime hair is fixed and at issue is how likely is it to match the given crime hair. Second, the hair may be quite peculiar and hence a match could be much more surprising. Third, there may be a number of hairs which may be matched. These issues emphasize the importance of the real or conceptual experiments providing the basis of the quoted sensitivity or FAR figures.

A further error is to erroneously use probability arguments or to assert figures which have no basis. It may be opined that in a certain city 10% of the population is black. Figures like this must be properly supported by evidence. Alternatively, basic probabilities are often multiplied to arrive at joint probabilities of occurrence when there is no basis for assuming that the associated attributes are independent. Sensitivity or FAR figures arrived at like this are worthless.

References

- Eggleston, Richard, Sir (1983). Evidence, proof, and probability (2d ed). Weidenfeld & Nicholson, London. Gaudette, B.D. (1978). Some further thoughts on probabilities and human hair comparisons. Journal of Forensic Sciences 23:758-763.
- Gaudette, B.D. and Keeping, E.S. (1974). An attempt at determining probabilities in human scalp hair comparison. Journal of Forensic Sciences 19:599-605.

On identification by probability

Russell V. Lenth
University of Iowa

The presentation by R.V. Lenth was subsequently published in the Journal of the Forensic Science Society 1986, Vol. 26, pages 197-213. It appears here with the permission of the Forensic Science Society.

ORIGINAL PAPERS

© Forensic Science Society 1986

On Identification by Probability

RV LENTH

*Department of Statistics and Actuarial Science, The University of Iowa,
Iowa City, Iowa 52242, USA*

Abstract

This paper considers the use of probability in identifying the source (such as a victim, criminal, or object) of some physical evidence (such as a bloodstain). It is common to use the probability that the characteristics observed in the evidence would occur at random; however, this is misleading because the number of possible sources of the evidence is not considered. When the corresponding attributes of the suspected source are not all known, one might use "non-exclusion" probabilities or related calculations; but these make inefficient use of the available information. Methods are given for taking the number of possible identities into account in both the known- and unknown-attributes cases. Finally, we consider the situation where partial information is available concerning several suspected sources, and suggest ways of explaining the calculations in layman's terms. *Key Words:* Identification; Probability; Associative evidence evaluation; Interpretation of bloodstain evidence.

Journal of the Forensic Science Society 1986; 26: 197-213
Revised version received 5 December 1985

Introduction

The purpose of this paper is to clarify some of the issues related to the use of probabilities in the identification of human beings or objects. Examples of situations in which such techniques could be used are numerous. One might wish to assess the chances that a bloodstain or dismembered body part belongs to a particular missing person; or that a bullet was fired by a particular weapon; or (based on blood typing) that a certain man is the father of a child; or that a suspect is the perpetrator of a crime, given that he has left behind some trace evidence that could have originated from him.

All of these examples are conceptually the same; certain evidence is available, and a person or object having physical characteristics that are consistent with the evidence has been identified. (For convenience, the term "suspected source" is used here to denote this person or object.) The objective is to determine how likely it is that the suspected source is the actual source of the evidence. Typically, the evidence for identifying a person consists of such things as race and sex, as well as the results of laboratory tests on tissue or bloodstains. Assume that information is available concerning the incidence of these characteristics in the population under study. The basic idea behind using probability in such situations is

clear; if a rare characteristic is observed, and if the suspected source has that same rare characteristic, then this constitutes stronger evidence in favor of the possibility that the suspected source is indeed the source than if no rare characteristics are observed. In practice, however, there is a great deal of confusion surrounding the choice of an appropriate probability to calculate what methods are correct for the calculation of these probabilities, and the interpretation of the results.

The simplest problem to consider is the case where we know whether or not the suspected source matches the evidence. If he (or she or it) does not, then we know with certainty that he is not the source of evidence. (This presumes that there are no errors in the evaluation—measurements, blood typing, and so forth—of the evidence. This assumption is made throughout this paper.) But if the suspected source matches the evidence, then we know nothing with certainty, and so we can express our degree of belief as a probability. There is an interesting history of such cases, the most notorious of which is *People v. Collins* [1]. The *Collins* case and others are discussed in Cullison [2]; the basic thrust of that paper (and the present one as well) is that the number of possible sources must be taken into account—otherwise the probability calculations can be misleading. The pioneering work in this area are two papers by Kingston [3, 4]. A subtle error in [3] is corrected here.

A more complicated situation arises when it is not known whether certain characteristics of the suspected source match the evidence. For example, there are a large number of genetic markers that can be detected from a tissue sample or a bloodstain, but these markers (excepting blood type) are seldom evaluated in the course of routine medical care, and hence they are unknown for most individuals. In such cases, information can be gathered by analyzing the characteristics of the suspect's family, such as is done in paternity testing. Stoney [5] discusses the relative merits of some *non-exclusion probabilities*, and argues that a certain *likelihood ratio* is more appropriate. That assessment is correct, but the likelihood ratio fails to take the size of the target population into account. This paper extends Stoney's results accordingly.

When statistical evidence is presented to a jury, it is important that it be as understandable as possible. Almost any probability used for identification, no matter how carefully it is explained, is likely to be interpreted by the layman as "the probability that so-and-so did it" (or the equivalent statement for the situation at hand). With this premise in mind, most methods in common use are misleading. This paper focuses upon methods that can directly address this natural tendency in interpreting probabilities, and upon ways of explaining the calculations in an understandable way. It is important to keep in mind that the issue of concern here is *identification* of the source of the evidence; the methods and formulas apply specifically to identification, and not to any other concept of association between a suspected source and a crime. Indeed, the use of probability calculations of any kind may be misleading in establishing some other kind of association.

In this presentation, it is assumed that each piece of evidence is of a categorical nature (for example, sex or blood type), as opposed to a measurement having a large number of possible values (for example, the length of a bone). In the case where some of the data consist of measurements, one could simply divide the set of all possible values of the measurement into classes and proceed as though it were a categorical variable. Alternatively, some methods for handling this kind of data are given in Lindley [6].

Case where the suspected source's attributes are known

Consider the case where, for each piece of evidence observed, we know with certainty that the suspected source has that same attribute. In *People v. Collins* [1], for example, part of the evidence was that a robbery was committed by a black man with a moustache and a beard and a blonde woman with a ponytail, and that the getaway car was yellow with an off-white top. A married couple was found who had all of these attributes, and they were accused of the crime. In presenting the case to the jury, the

prosecutor argued that this particular combination of attributes is extremely rare, and hence the odds are heavily in favor of the accused couple being the actual perpetrators of the robbery. His "expert" witness presented the characteristics and probabilities given in Table 1 in support of this argument. He then multiplied these probabilities together to obtain a probability of 1/12,000,000 for the joint incidence of the attributes, and

TABLE 1. Probabilities used in the Collins case

Characteristic	Probability
Partly yellow automobile	1/10
Man with moustache	1/4
Girl with ponytail	1/10
Girl with blonde hair	1/3
Negro man with beard	1/10
Interracial couple in car	1/1000

concluded that there was only one chance in twelve million that the defendants were innocent.

This example is used because it illustrates practically every error that can be made in solving a problem of this type. The three most major errors are the following: there is no factual basis for the probabilities; the characteristics are not independent; and even if the numbers were correct, the interpretation given to the joint probability of 1/12,000,000 is incorrect. The first two points are not discussed here; the reader is referred to Cullison [2]. We concentrate instead on the third, the interpretation of joint incidence probabilities.

While mathematical notation is being avoided as much as possible, a little bit is useful. Let the letter p denote the joint probability of incidence of all of the attributes observed in the evidence. In other words, a randomly drawn person (or in the Collins case, couple) would have a probability of p of matching all of the characteristics. For purposes of exposition, pretend that the probability calculations in the Collins case are correct, and hence $p = 1/12,000,000$ in this example. Let the letter S denote the event that the suspected source (in this case, a couple) is indeed the source of the evidence.

The goal is to evaluate $P(S)$, the probability of the event S . Actually, this problem is incorrectly stated, because S is not random. Either S is true or it is not, and so $P(S)$ is either one or zero. This ambiguity can be dismissed by attaching a slightly different meaning to $P(S)$. If it were possible to replicate a crime having outwardly identical circumstances a very large number of times, then $P(S)$ is defined to be the proportion, in the long run, of cases in which the suspected source is the actual source of the evidence. The following analogy might help clarify this issue. Before a horse race, it makes sense to talk about probabilities because the outcome is uncertain. After the race has been run, there is no longer any uncertainty as to the facts, but there remains uncertainty in one's knowledge of the outcome unless he has been told the results. In such a situation, it makes sense to use the probabilities that pertained before the race was run in order to express one's beliefs concerning the outcome.

There is a tendency for the layman to interpret practically any probability that is presented in connection with an identification problem as either $P(S)$ or $1 - P(S)$ (depending on the context), no matter how carefully it is explained. In the Collins case, for example, the value $p = 1/12,000,000$ was explicitly presented as the value of $1 - P(S)$, so that the jury was led to believe that $P(S) = 11,999,999/12,000,000$. In general, it is a common mistake to state or imply that the incidence probability p is the same as $1 - P(S)$, or equivalently that $P(S) = 1 - p$. This statement is, in fact, almost never correct. The reason is that the size of the target population (that is, the number of possible sources of the evidence) is not taken into account; the more possible sources there are, the less compelling a small value of p becomes. The incidence probability $p = 1/12,000,000$ (which we are pretending is correct) is quite convincing when there is only a handful of couples who could actually be the robbers; but if there were, say,

36,000,000 other couples besides the accused who, before the evidence is considered, could equally well be the robbers, then we would expect about three of them (1/12,000,000 of 36,000,000) to match the evidence. In this case, the evidence against the accused couple is far from compelling.

Some methods for taking the population size into consideration are given in Kingston [3]. One method uses a probability model that assumes that the true origin of the evidence has been chosen at random from the set of all people having the same attributes as the evidence. In order to explain Kingston's method, we need the notion of *conditional probability*. The conditional probability of A , given B , denoted by $P(A|B)$, is the probability of A occurring when we restrict our attention to cases where the event B can occur.

In the identification problem, we often do not know exactly how many people (or objects), x , actually match the characteristics of the evidence, and this makes it more difficult to compute $P(S)$; but it is easy to compute the conditional probabilities $P(S|x)$ for any x . If $x=1$, then we have identified the only possible source, and hence $P(S|1)=1$. For $x=2$, we have identified one of the two possible sources, so that under Kingston's randomness assumption, $P(S|2)=1/2$. In general, we see that $P(S|x)=1/x$ for each whole number x . The unconditional probability of S , $P(S)$, can then be found by computing a weighted sum of the values of $P(S|x)$, where the weights are the probabilities of the individual values of x :

$$\begin{aligned} P(S) &= P(S|1)P(1) + P(S|2)P(2) + P(S|3)P(3) + \dots \\ &= P(1) + P(2)/2 + P(3)/3 + \dots \end{aligned}$$

Kingston notes that if there are n possible sources, and the probability of the evidence is p , then the value of x is like the number of "heads" one might observe if a biased coin having $P(\text{heads})=p$ is flipped n times independently; this is called the *binomial* probability model. He also argues that the case $x=0$ is impossible, since there is at least one person (namely, the true source of the evidence) who matches the evidence. Thus, the values of $P(x)$ are obtained from the binomial probability distribution after conditioning on the fact that x is not equal to zero. In the case where np is a small number, this leads to the formula:

$$P(S) = 1 - np/4 \quad (\text{approximately}).$$

The case where np is not small (say, greater than 0.2) is not particularly relevant, since in this case the evidence is not very compelling. For illustration, suppose in the *Collins* case that there are $n=4800$ possible couples (including the defendants) who could be the robbers, and that $p=1/12,000,000$; then $np=4800/12,000,000=0.0004$, and so Kingston's rule says that $P(S)$ is approximately equal to $1 - 0.0004/4 = 0.9999$. On the other hand, if there are 36,000,001 possibilities, as in the earlier illustration, then $np=3$ approximately. This is not a small number, and in lieu of applying the formula, we would say that the evidence considered in these calculations gives us no compelling reason to believe that the defendants are guilty.

Recall that the model used in the above formula assumes that the alleged source of the evidence is a random selection from those persons having the required characteristics. If we are unwilling to believe this, then the conditional probabilities $P(S|x)$ cannot be used. The conservative thing to do is to simply report the probability that x is equal to one, that is, that there are no other sources of the evidence other than the suspected source:

$$P(\text{Unique}) = P(1) = np(1-p)^{n-1}/[1 - (1-p)^n].$$

This is approximately $1 - np/2$ when np is small. For any sampling model, $P(S)$ is always greater than or equal to $P(\text{Unique})$; that is the meaning of the term "conservative" in this context.

Kingston's approach includes the alleged source of the evidence among the n persons under question, and then conditions upon the fact that there is at least one matching person in this collection. An alternative way to look at the problem is to work in terms of the number y of matching sources besides the alleged one. Clearly, $P(S|y) = P(S|x-1)$ since y is always one less than x , and the probability distribution for y is unconditionally binomial

where now there are only $n - 1$ flips of the biased coin. Denoting the probabilities for the y values by $Q(y)$, then under the randomness model,

$$\begin{aligned} P(S) &= Q(0) + Q(1)/2 + Q(2)/3 + \dots \\ &= [1 - (1 - p)^n]/np \\ &= 1 - (n - 1)p/2 \quad (\text{approximately}). \end{aligned}$$

Using the approach where no randomness is assumed, we have:

$$P(\text{Unique}) = Q(0) = (1 - p)^{n-1},$$

which is approximately equal to $1 - (n - 1)p$ when np is small. This is a disturbing discrepancy because, while it appears that we should have gotten the same results, the values of $1 - P(S)$ are about twice as large in both formulas when we work in terms of y instead of x .

The reason for this aberration is quite subtle, but it can be understood by considering a simple case. Suppose that $n = 2$; that is, there are two possible sources for the evidence. One of them is the actual source (denoted S); the other one is not the source (NS). Using the notations M for matching and NM for not matching the evidence, there are four ways in which the symbols can be combined:

Case 1. $S:M$, $NS:M$ Case 3. $S:M$, $NS:NM$
Case 2. $S:NM$, $NS:M$ Case 4. $S:NM$, $NS:NM$

Kingston conditions on x being nonzero, that is, at least one M appears on the list; thus, case 4 is excluded. But case 2 is not excluded in Kingston's rule, even though it is impossible. (The source of the evidence *must* match the evidence, because we are assuming that the measurements are correct, and that the attributes under consideration remain constant.) The second approach (using y) excludes both case 2 and case 4; Case 1 is the same as $y = 1$, and case 3 is the same as $y = 0$, since this is the only way that only one match can exist. This is yet another example of just how tricky problems of this type can be.

Now consider a situation in which the source of the evidence might not match the evidence—due to imperfect measurements, for example. Kingston's rules do not apply here either. Having imperfect measurements certainly does not give us the right to cut in half our stated chance of making an incorrect identification! Instead, we must consider a much larger pool of possible sources for the evidence, namely those who could conceivably agree with the evidence based on the same measurement procedure as was used for the evidence.

To summarize, the corrected versions of Kingston's rules are:

$$\begin{aligned} P(S) &= 1 - (n - 1)p/2 \quad (\text{approximately}), \\ P(\text{Unique}) &= 1 - (n - 1)p \quad (\text{approximately}). \end{aligned}$$

To illustrate these in the *Collins* case, again consider $n = 4800$ and $p = 1/12,000,000$. Under the randomness model, the approximate value of $P(S)$ is $1 - 0.0004/2 = 0.9998$. The probability of uniqueness of the defendants' attributes is approximately $1 - 0.0004 = 0.9996$. The exact values of these probabilities are identical to the approximations to six decimal places.

Case where the suspected source's attributes are unknown

It is often true that we do not know whether all of the characteristics of the evidence hold for the suspected source. For example, several genetic markers can be measured by analyzing a bloodstain or a tissue sample, but these markers are not usually observed in the course of routine medical care, and hence it would not be known for certain whether or not a suspect source has all of the same phenotypes as are found in the bloodstain.* However, information can often be gained by analyzing samples from the suspected source's family.

* For the uninitiated, a *phenotype* is an observable trait. A *genotype* is a genetic trait; it may or may not be observable.

An example of this type is given in Kuo [7] and it is discussed further in Stoney [5]. Data are given concerning six phenotypes observed in a bloodstain and in a missing person's parents; they are displayed in Table 2.

TABLE 2. Bloodstain evidence in Kuo example

Marker	Stain	Father	Mother
ABO	B	B	O
EAP	A	BA	BA
AK	2-1	1	2-1
ADA	1	1	1
PGM	2-1	2-1	1
Hp	2-1	2-1	2

Stoney discusses the relative merits of two probabilities that might be used: *non-excluded couples* and *non-excludable stains*. He then presents a better measure, called the *likelihood ratio*. They are briefly described here, and some comments are added concerning the interpretation of the likelihood ratio. All undocumented computations in what follows are given in detail in Stoney's paper.

The *frequency* (or probability) of *non-excluded couples* is the fraction of couples in the population who could not be excluded as possible parents of a person having the observed phenotypes. In the example, non-exclusion probabilities can be calculated separately for each of the six systems using standard genetic theory. It is reasonable to assume that the systems are independent (see Grunbaum and colleagues [8]), so the combined non-exclusion probability is obtained by multiplying the results together, yielding a value of 0.0062. Note that the couple's phenotypes are not used in this calculation; that information is used only in interpreting the non-exclusion probability, where it is simply noted that the couple's phenotypes are consistent with those found in the stain.

The *frequency of non-excludable stains* is obtained by reversing the roles of the couple and the stain. We find the probability that a randomly chosen stain is a possible result of combining the phenotypes of the couple. Again, separate calculations are made for each system and we multiply the results together (by virtue of independence). The resulting non-excludable stain probability is 0.3386. Note that the phenotypes in the stain are not used in the calculations.

Of course, one could also compute the *phenotypic incidence* of the stain, this being the probability p that the observed combination of phenotypes would occur at random, just as in the preceding section of this paper. In the Kuo example, the phenotypic incidence is $p = 0.00012656$. Since it is based on the stain phenotypes rather than the couple's, it is more closely related to the probability of a non-excluded couple than to that of a non-excludable stain. But it is not the same as the non-excluded couple probability. The incidence probability is much smaller than the probability of a non-excluded couple. In fact, the non-exclusion of couples probability is always at least as large as the phenotypic incidence. The rules (Kingston's rules as corrected) described above for the interpretation of the phenotypic incidence do not apply here, because it is not known for certain whether the traits of the suspected source match those of the stain.

Which one of these probabilities is most meaningful? Stoney [5] argues that we should proceed from the known to the unknown; since the couple's phenotypes are known while the person whose bloodstain we have observed is unknown, the probability of non-excludable stains should be used. He also notes that if the couple is completely homozygous (*i.e.*, the partners have the same genotypes), then the frequency of non-excludable stains reduces to the phenotypic incidence p , while at the same time the suspected source's phenotypes would be known completely; this would put us back into the situation of the case where the suspected source's attributes are known. The frequency of non-excluded couples does not share this property, and this is a valid argument against using it as a meaningful measure. On the other hand, the frequency of non-excludable stains does

not use the main piece of evidence (the bloodstain)—other than to simply observe whether it is consistent with the couple's traits. If the bloodstain is a rare combination of phenotypes, then that is helpful information in identifying the source, and it ought to be used. By the same token, if the parents of the suspected source can only produce a very limited variety of phenotypes in their offspring, that again is useful information. The failure of both non-exclusion probabilities is that they don't consider the couple and the stain *simultaneously*.

The *likelihood ratio* addresses this question. It is the ratio of two conditional probabilities of the stain phenotypes:

$$LR = P(\text{stain} \mid \text{couple}) / P(\text{stain} \mid \text{some other couple}) = p' / p.$$

Here, p' is the probability that the parents would pass the observed combination of phenotypes on to their offspring, while p is the phenotypic incidence. If the couple are not the parents of the source, then we know nothing about the parents of the source and hence we act as though they were randomly selected. Note that the couple is being used as the source of information about the suspected source, and the stain is the evidence at hand; thus the likelihood ratio can be more generally written as follows:

$$LR = P(\text{evidence} \mid S) / P(\text{evidence} \mid \text{not } S),$$

where S has the same meaning as in the case where the suspected source's attributes are known. In the bloodstain example, the phenotypic incidence of the stain is $p = 0.00012656$, and the probability that the stain phenotype would occur in the couple's offspring is $p' = 0.0164$. Thus, the likelihood ratio is $LR = 0.0164 / 0.00012656 = 130$. The suspected source is 130 times more likely to have the observed phenotype than a randomly selected person.

Note that if all of the attributes of the suspected source are known and match the evidence, then $P(\text{evidence} \mid S) = 1$, and so $LR = 1/p$. In general, the reciprocal, $1/LR$, of the likelihood ratio is like an incidence probability that has been adjusted for the amount of uncertainty concerning the suspected source's phenotypes. Using it as though it were $P(S)$ is as erroneous as before. Stoney issues the same warning: "In no sense [does it] . . . represent a likelihood or a probability that the accused was involved in the crime. In contrast, [it] represent[s] the likelihood or probability that the incriminating aspects of the evidence would occur if the accused is uninvolved." It seems dubious, though, that this message would be well understood by a jury, even after repeated restatement of this warning.

The likelihood ratio can be used to derive generalized versions of Kingston's formulas for $P(S)$ and for $P(\text{Uniquely } S)$. (The term "uniquely S " must be used now instead of "unique," because it is possible that there is only one person having the traits, but it is not the suspected source.) Recall that y is the number of people *other than the suspected source* who match the evidence. Now let z be the number of *suspected sources* who match the evidence; that is, z is either one or zero depending on whether or not the suspected source has the same attributes as the evidence. Note that $x = y + z$. Note also that $z = 1$ has probability p' and $z = 0$ has probability $1 - p'$. The case $\{y = 0, z = 0\}$ is impossible because there is at least one match for the evidence in existence. Using this information, it can be shown that Kingston's formulas (as corrected) can be generalized to this situation by simply applying a correction factor CF:

$$P(S) = CF[1 - (n - 1)p/2] \quad (\text{approximately}),$$

$$P(\text{Uniquely } s) = P(y = 0, z = 1) = CF[1 - (n - 1)p] \quad (\text{approximately}),$$

where

$$CF = p' / [1 - (1 - p')(1 - p)^{n-1}].$$

Note that when $p' = 1$, then $CF = 1$, and both $P(S)$ and $P(\text{Uniquely } S)$ reduce to the formulas we use when the suspected source's attributes are known. Thus, they can be used in all cases, as long as p' is considerably larger than p (when this condition is not satisfied, the evidence in favor of the suspected source is not compelling).

The above results can be derived by observing that y and z have the distribution of independent binomial random variables, except it is truncated to exclude the case where they are both zero. Thus, each remaining probability must be divided by $1 - P(y=0, z=0) = 1 - (1-p')(1-p)^{n-1}$ (the denominator of CF). The formulas given are first-order Taylor series expansions (as are the approximations given earlier) of the appropriate probabilities.

For illustration, suppose in the Kuo example that there are 25 missing persons (the suspected source and 24 others) who cannot be excluded as possible sources of the bloodstain. Then the correction factor is:

$$\begin{aligned} \text{CF} &= 0.0164/[1 - (1 - 0.0164)(1 - 0.00012656)^{24}] \\ &= 0.0164/[1 - 0.9836(0.99987344)^{24}] \\ &= 0.8467. \end{aligned}$$

The probabilities of interest are thus:

$$\begin{aligned} P(S) &= 0.8467[1 - 24(0.00012656)/2] = 0.8454; \\ P(\text{Uniquely } S) &= 0.8467[1 - 24(0.00012656)] = 0.3441. \end{aligned}$$

Recall that when the suspected source's traits are all known, $1 - P(S)$ is about twice as large as $1 - P(\text{Uniquely } S)$. In the present example, they hardly differ. This is because p is extremely small and so the correction factor dominates the calculations. The two probabilities become more disparate if either p or p' is increased.

When n is large and p is small, one might incur significant roundoff errors in computing $(1-p)^{n-1}$; in this case we can use the fact that

$$(1-p)^{n-1} = \exp[-(n-1)p] \quad (\text{approximately}),$$

where $\exp[x] = e^x$ is the exponential function. This is also somewhat easier to compute. In the example, the approximate correction factor is very close to the exact value:

$$\begin{aligned} \text{Approx. CF} &= p' / \{1 - (1-p')\exp[-(n-1)p]\} \\ &= 0.0164 / \{1 - (1 - 0.0164)\exp[-24(0.00012656)]\} \\ &= 0.8461. \end{aligned}$$

Kingston [4] suggests a Bayesian approach to the problem. A theorem in probability known as Bayes's rule can be used to reverse the conditions in a conditional probability. In this case, we know $P(\text{evidence} | S) = p'$ and $P(\text{evidence} | \text{not } S) = p$; we wish to find $P(S | \text{evidence})$. (This is the same as $P(S)$ in the discussion above; the conditioning upon the evidence was implicitly understood.) An informal development of the answer is as follows. If there were one person besides the suspected source who could be the source of the evidence, then the betting odds in favor of the suspected source would be $p' : p$ (the likelihood ratio). It thus seems logical that if there are $n-1$ possible sources besides the suspected source, the odds would be $p' : (n-1)p$. This translates to a probability of

$$\begin{aligned} P(S | \text{evidence}) &= p' / [p' + (n-1)p] \\ &= 1 / [1 + (n-1)p/p']. \end{aligned}$$

It is interesting to note that this is approximately equal to $1 - (n-1)p/p'$ when p/p' is small; this is the same as our approximation to $P(\text{Uniquely } S)$. $P(S | \text{evidence})$ is always smaller than $P(S)$ using the generalization of Kingston's rule, because it does not incorporate any assumptions regarding the manner in which the suspected source was chosen. The simplicity of the odds argument is an attractive feature of the Bayesian approach; it can be explained in simple terms.

In the Kuo example, recall that the likelihood ratio is 130. As before, suppose that there are 24 persons besides the suspected source who are possible sources of the bloodstain. Then the Bayesian approach is as follows: the odds in favor of the suspected source against each one of the others are 130 : 1, so the combined odds against all 24 others are 130 : 24. Thus, the probability that the bloodstain comes from the suspected source is $130/(130 + 24) = 0.8442$. Note that this is almost identical to the values we obtained from the generalizations of Kingston's rules. The

Bayesian method can also be used in the case where the attributes of the suspected source are known. We simply use $p' = 1$, or equivalently, $LR = 1/p$. In the Collins example considered above $LR = 12,000,000$, so that if $n = 4800$, the odds of the suspected source are $12,000,000 : 4799$ for a probability of $12,000,000 / (12,000,000 + 4799) = 0.9996$. This is indistinguishable from the uniqueness probability calculated earlier.

Incorporating more specific information

It is possible that, in the context of the preceding section, some information in addition to genetic markers is available. For example, by examining human remains it might be possible to determine such things as sex, race, and approximate age: this information should be incorporated in the analysis. Furthermore, while some attributes of the suspected source might be unknown, other characteristics might be known. A person's blood type, for instance, is often available while some other genetic markers are not known; we should take advantage of the available information. In fact, in the course of investigation we are likely to gather partial information about persons other than the suspected source. Some possible sources are eliminated on the basis of this information, thereby reducing the value of n and increasing $P(S | \text{evidence})$. But we might also find, for example, that certain alternative sources have the same blood type as the evidence sample. This has the effect of increasing the value of p for those persons, thereby decreasing $P(S | \text{evidence})$. Of course, if this information is not included in the calculations, then the value of $P(S)$ is larger than it should be, an especially dangerous circumstance.

The Bayesian approach is especially attractive in this situation, as is presently demonstrated. It can easily be generalized to the case where each possible source has a different probability of matching the evidence. This allows as much information as possible to be taken into account, so long as it meets the important criteria that objective information regarding the probability of each attribute must be available; and that it must be reasonable to assume that the attributes occur independently of one another, or if they are dependent, that appropriate conditional probabilities are available.

An illustrative example is the case of *State v. Klindt* [9]. The case involves the identity of a portion of a woman's body. By analyzing the remains, it was found that the woman was white, between the ages of 27 and 40, had given birth to at least one child, and had not been surgically sterilized. In addition, seven genetic markers were identified. On the basis of this and other evidence, it was possible to eliminate all but four missing persons (label them W , X , Y , and Z) in the four-state area where the body was found, as possible identities. Woman W had been missing for one month at the time that the body was discovered, and she had last been seen in the same area. The other three had been missing for six months, six years, and seven years, respectively, and their last known locations were all at least 200 miles from where the body was found.

We know that W 's blood type is A, and hence that phenotype is treated separately from the others in the calculations (the other markers are unknown for all four women). For the remaining six phenotypes, tissue samples from W 's parents were used to calculate $P(\text{phenotypes} | W) = 0.5$, compared to a phenotypic incidence of 0.00764 for the population as a whole. No familial testing was done for the other three possible victims. The ages of all four women are known, and they are all known to be mothers; however, the blood types of X , Y , and Z , are unknown. For the sake of illustration, though, the following alterations were made to the actual data: it is not known whether or nor X is a mother; Y 's age is unknown; and Z is known to have type A blood.

The probabilities are summarized in Table 3. Anywhere that an entry of 1.0 appears, it corresponds to a case in which a trait is known to hold for that person. A zero entry would indicate that that trait is absent (all such cases can be eliminated as contenders, and hence they do not appear in the

table). The numbers between zero and one indicate cases where the presence or absence of the attribute is unknown; the number given is the incidence probability of that trait.

TABLE 3. Probability evidence in the Klindt case (altered slightly for illustration purposes)

	Women			
	W	X	Y	Z
Age (years)	33	27	unknown	37
Trait				
Mother Age	1.0	0.583	1.0	1.0
Not sterilized Mother, Age	1.0	0.839	0.662	0.542
Type A blood	1.0	0.362	0.362	1.0
Other six phenotypes	0.500	0.00764	0.00764	0.00764
Likelihood	0.500	0.00135	0.00183	0.00414
Fraction of total	0.986	0.0027	0.0036	0.0082

Let us examine Table 3 row-by-row. The entries for age are self-explanatory; they are important here because the incidences of motherhood and surgical sterility are age-dependent (older women are more likely to have given birth and to have been sterilized). This is reflected in the "Mother | Age" row for person X (the others are known to be mothers). In addition to its dependence on age, the incidence of surgical sterility is quite different for women who have borne children and those who have not. For women X and Z, age-specific values are given, while the probability of being not surgically sterile for person Y is an average of the age-specific probabilities for ages 27 through 40; this average is weighted by the numbers of women in the U.S. population at each age. Note that age, parity, and sterility are not independent, but we can still use the information because the appropriate conditional probabilities are available. It is reasonable to assume that blood type and the other six genetic markers are independent of the other factors and of one another.

For all of these traits, the most race- and sex-specific information available is used. The probabilities pertaining to incidence of motherhood and surgical sterility are obtained from the population of white women, and those pertaining to the genetic markers apply to white persons—no sex-specific information was available. The data come from various sources, including the 1980 U.S. Census, the Public Health Service, and Grunbaum and colleagues [8].

The *likelihood* is the joint probability of all of the traits for each person. It is obtained by multiplying together all the entries (except, of course, age) in each column. Each likelihood is in fact $P(\text{evidence} | \text{person})$. Bayes's theorem can be used to find the values of $P(\text{person} | \text{evidence})$. If we assume that, aside from the information presented in the table, each person is equally likely to be the identity of the body, then these probabilities are obtained by simply dividing each likelihood into the total of the likelihoods. These numbers are given in the "fraction of total" row. We see that $P(W | \text{evidence})$ is 0.986, indicating that the body is very likely to be W.

This can be easily explained in layman's terms. View each row of the table as a screen through which the information concerning the possible sources' attributes is sifted. Starting, for example, with 1000 lb of "sand" having person X's attributes, then only 583 lb remains after sifting it through the "mother" screen; 83.9 percent of the surviving sand, or 489 lb, makes it through the "non-sterility" screen, and so forth; 1.35 lb of sand remains after sifting it through all of X's screens. By comparison, all of W's metaphorical sand survives the first three screens, but half of it is eliminated in the final screen, leaving us with 500 lb. Similarly, the weights of sand for Y and Z are 1.83 and 4.14 lb respectively. The total amount of sifted sand for the four women is about 507 lb. These amounts of metaphorical sand correspond to the weights of evidence in favor of each of the women being the actual victim. The fraction of sand attributable to woman W is $500/507 = 0.986$, and thus she is by far the most likely identity of the victim.

The presenter of the above analysis is likely to be confronted with the following rebuttal: "The reason that the likelihood (or amount of sand) for W is so much higher than for X , Y , or Z is that much more information has been gathered about her than the three (as evidenced by the fact that there are so many 1's in that column of the table); thus, the results are a foregone conclusion." Although it may seem convincing at first, this argument is not valid for two reasons. The main reason that X , Y , and Z appear in the table at all is that little is known about them; if more information were available, we probably could have excluded one or more of them from consideration. In spite of the large amount of available information about W , none of it is contrary to the evidence. The common-sense meaning of this is that you can identify a close friend reliably without collecting information about the traits of other people whom you don't know.

It is necessary to examine the shortcomings of an analysis like this one. Defects and inadequacies can sway the results in either direction, but it would be most desirable if they make the results demonstrably conservative; that is, we have underestimated $P(W | \text{evidence})$. In the Klindt example, some minor errors may result from the fact that the wrong population is used to obtain the probabilities. It is unlikely that the fraction of mothers or surgically sterile women is exactly the same among missing white women as it is among white women in general. But the values used are the best available, and they do not bias the results extensively. The analysis neglects the possibility (which is unlikely) that the body could be from outside the four-state area considered. A more serious defect is the fact that the body could be that of a woman not on the missing-person list, and hence someone other than the four women considered.

While these inadequacies might inflate the value of $P(W | \text{evidence})$ unrealistically, the effect is more than offset by the implicit assumption that, aside from the factors considered in the analysis, the four women are equally likely to be the victim. There are other factors that could not be used because no objective data are available. The primary ones are the location of the body and the times of the women's disappearances. If the information existed, it would contribute more rows to the table and alter the likelihood values. While it is hazardous to guess how much this would affect the numerical results, it is safe to say that it makes the calculations conservative. By carefully explaining the fact that other evidence, if it were quantifiable, would simply add more rows to the table and alter the likelihoods accordingly, then a jury can more fully understand how the probability calculations fit in with the rest of the case.

Conclusions

This paper attempts to expose the ways in which probabilities can be used, both properly and improperly, in the identification of a victim, a criminal, or an object. It is a natural inclination to view almost any probability as a probability of correct identification, even when extreme care is taken in explaining its meaning. Incidence probabilities, probabilities of non-excluded couples or stains, and likelihood ratios are inadequate for that reason.

In order to avoid misleading results, the number of possible sources must be included in the analysis. Kingston's formulas, as corrected and generalized in this paper, may be used for this purpose; these put the incidence probability or likelihood ratio into proper perspective with regard to population size. However, Bayes's theorem has the additional advantage of being easily explained (provided that we avoid the approach used in most elementary probability textbooks!). The Bayesian approach yields approximately the same results as Kingston's formula.

In realistic cases, there are three kinds of evidence: probabilistically quantifiable factors that are either known or unknown to hold in the suspected source, and nonquantifiable factors. Quantifiable factors can be combined efficiently using Bayes's theorem. The calculations can be explained in concrete terms by using the metaphor of sifting sand through a succession of screens. The manner in which nonquantifiable factors would enter into the calculations if the information were available can also be explained; this better enables the layman to judge the importance of the statistical arguments in the overall case.

References

1. People v Collins. California Reporter 1968; 66: 497.
2. Cullison AD. Identification by probabilities and trial by arithmetic (a lesson for beginners in how to be wrong with greater precision). Houston Law Review 1969; 6: 47.
3. Kingston CR. Applications of probability theory in criminalistics. Journal of the American Statistical Association 1965; 60: 70.
4. Kingston CR. Applications of probability theory in criminalistics—II. Journal of the American Statistical Association 1965; 60: 1028.
5. Stoney DA. Statistics applicable to the inference of a victim's blood type from familial testing. Journal of the Forensic Science Society 1984; 24: 9.
6. Lindley DV. A problem in forensic science. Biometrika 1977; 64: 207.
7. Kuo M. Linking a bloodstain to a missing person by genetic inheritance. Journal of Forensic Sciences 1982; 27: 438.
8. Grunbaum DW, Selvin S, Pace N and Black DM. Distribution of gene frequencies and discrimination probabilities of 22 human blood genetic systems in four racial groups. Journal of Forensic Sciences 1980; 25: 428.
9. State v Klindt. District Court of Scott County, Iowa; Case Number 115422, 1968.

Quantification and modeling of criminal careers

Alfred Blumstein and John P. Lehoczky
Carnegie-Mellon University

John Lehoczky

Professor Blumstein and I are going to present a talk in an area of criminal justice research of interest to most of you: criminal career modeling and quantification. The presentation will be divided into two parts. First, Professor Blumstein will offer a broad high-level picture of the entire research area. This will include topics such as why study criminal careers, partitioning aggregate crime rates, the conceptualization of criminal careers, the data sources available, and incapacitation effects. Al is one of the most significant individuals in criminal justice research today. He is especially noted for bringing operations research, probability modeling, and statistical analysis methodology to the study of many aspects of the criminal justice system. He is a major national figure in incapacitation policy and chairman of a National Academy of Sciences Panel on Criminal Careers. This panel will present its report in late 1986.

After this broad overview on criminal careers research, I am going to discuss the probability modeling of criminal careers that has been done, some of the drawbacks in it, and some of the ways in which it can be improved. In addition, I have several specific issues that I want to present. Some may be well known to a few of you in the audience, but I think they are very important. One is using models to correct various kinds of biases that arise in most criminal justice data sets. These will include two versions of the length-biased sampling problem. Second, I will present material on hierarchical modeling and the associated empirical Bayes estimation techniques. The handout set has extensive references at the end for your information. These references offer a broad survey of the whole spectrum of research. I will now turn the floor over to Al Blumstein.

Alfred Blumstein

The issue of criminal careers is one that has been around in criminology for quite a while. It was a major thrust of the work of the Gluecks in the early 1930's. (1) That perspective seemed to have faded for a while, but it is receiving increasing attention today.

Empirical literature in criminology that tries to address causes of crime is generally characterized by attempts to relate aggregate crime rate as a dependent variable to a wide variety of exogenous co-variates. The basic dependent variable then has been aggregate crime rates. People have typically used cross-

sectional data, with states or cities as the units of observation, or time series data, most often using aggregate national crime rates. Part of the problem with aggregate crime rate as an indicator of crimes per population is reflected by the fact that it is equally good as an indicator of crimes per victim or crimes per offender. There are many ecological correlates that may be associated with victims or with offenders, and any causal research oriented at leading to a policy intervention should be able to distinguish intervention with victims from intervention with offenders. There must be a finer level of analysis than the very gross aggregate crime rates. One such involves use of the age-specific arrest rate; this begins to provide some information about who the offenders are.

Figure 1 is a graph of the 1983 age-specific rates as a function of age for the three index crimes of robbery, burglary, and aggravated assault. The graph shows a very sharp peak in the late teen years, with a rapid fall-off after the peak. The graph is very similar for all three crimes, with the peak occurring somewhat later (21 compared to 17) for aggravated assault.

The fall-off is very rapid in all three cases. It is fastest for burglary, reaching a rate half the peak in four years, by age 21; it is slowest for aggravated assault, which drops to the half-peak rate at age 36. Robbery is very close to burglary in its age patterns, reflecting the fact that—at least from the viewpoint of the offender—it is much more a property crime than the violent crimes with which it is normally grouped.

This aggregate data is reasonably well known, and has given rise to a variety of perceptions about criminal careers. One such view claims that because the rate at age 30 is so low, 30 year old offenders will soon be dropping out of their criminal careers, and so are poor candidates for incarceration. Furthermore, it is often argued from Figure 1 that offenders are most active when they are teenagers, and so they must slow down in their rate of crime commission at later points in their criminal career, and this again warrants paying less attention to older offenders. As I will point out, most such perceptions derived from aggregate data are deficient in various ways.

That takes us to the issue of why we want to study individual criminal careers, and why the aggregate data are insufficient. Let me first discuss the frequently transposed doublet, "career criminals" and "criminal careers." "Career criminals" refers to people who are at

the serious end of an offending scale. "Criminal careers" is simply a metaphor that uses the concept of a career to characterize a longitudinal sequence that has the basic elements of initiation, probably termination at some later time, and a variety of transitions in between.

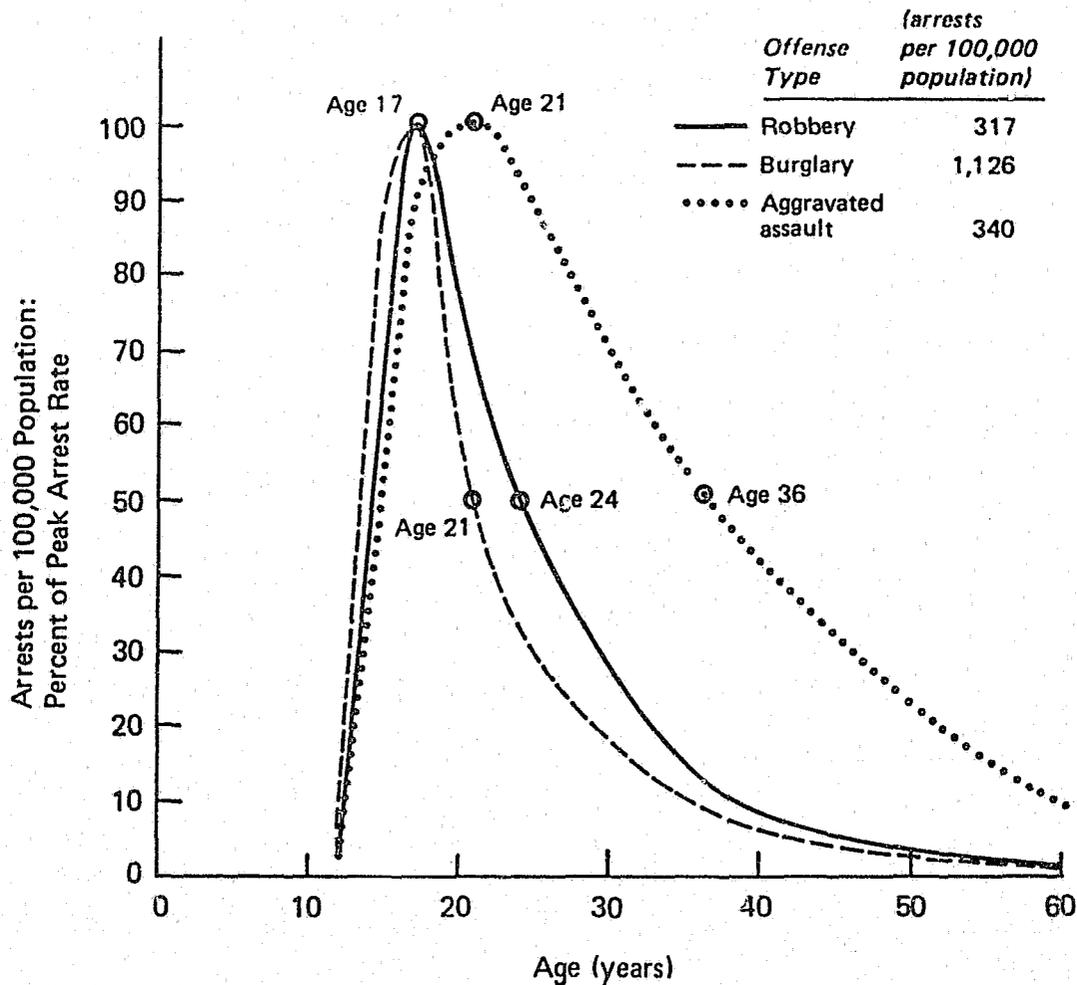
We want to focus on the criminal career as soon as we recognize that it is individuals who commit the crimes. They may do it in groups, but those are groups of individuals, and we want to be able to see the difference in what applies to individuals as distinct from the aggregate data.

Of course, the criminal justice system makes decisions about individuals, and the earlier illustrations indicated that it does try to take account of individual attributes. I plan to highlight some of the problems in making inferences from aggregate data to reach decisions about how to deal with an individual with a particular attribute like age.

Finally, because a criminal career is a dynamic longitudinal process, we would like to find ways for analyzing and characterizing that process. In moving from aggregate data to data on individual criminal careers, we first want to partition the aggregate rates into a rate of participation (i.e., that distinguishes those who are active participants in crime from those who are not), and then to focus on the criminal career parameters--especially the individual annual crime frequency--of those who are active.

There is a basic identity, $C = \lambda P$, linking the aggregate crime rate (C) to the number of criminals per capita in the population (P) and the annual frequency of crimes per criminal (λ). The prevalence, P , invokes the processes of initiation and termination; we are also interested in the interval between initiation and termination, or the career length. The identity indicates that the crime rate (C), the number of crimes per year per population, is simply the product of the annual frequency per active criminal (λ) and the prevalence of active criminals in the population (P).

FIGURE 1 Age-Specific Arrest Rates for Robbery, Burglary and Aggravated Assault.



Once these criminal-career parameters have been measured, it will then be desirable to explore each parameter's specific determinants. In doing so, it is important to recognize that the determinants of any single parameter may well be different from the determinants of any other parameter.

We can now go on to examine some features of criminal careers. A criminal career starts at some onset point. This might be the occasion of the first crime, or it might be the first arrest, or it might even be the occasion when the individual first has an elevated propensity to commit crimes. Once initiated, crimes then occur in some stochastic process at some rate that may be constant or may be varying over time. Eventually, the criminal career will terminate when the individual drops out of his criminal career. The "career length" is the interval between termination and initiation.

Figure 2 is the simplest possible representation of a criminal career, showing a constant rate of offending throughout the career, and instantaneous initiation and termination. It is possible, of course, that the frequency, λ , is not constant over the career (for example, it may display a trend or it may merely fluctuate), or that there is a finite rise time over which the value of the individual's frequency climbs to a constant value.

If one were to imprison that individual for a period S during the criminal career, then removing that individual from the community might potentially avert λS crimes, as depicted in Figure 3. Realizing that potential requires that the offender take his crimes off the street with him. That is, if they represented personal violence or sociopathic activity, they can be expected to leave with him. If, however, he were selling drugs as part of a drug ring, the crimes would stay on the street because a criminal labor market would generate someone else to replace him. Thus, by no means do all crimes go to prison with the prisoner.

In addition, some prisoners might bring their crimes into prison with them, and might even commit more crimes in prison than on the outside. Understandably, however, the society somewhat discounts the crimes in prison compared to crimes in the community.

The consequences of the imprisonment could also show itself as a change in the prisoner's career after release. This could be a rehabilitative effect, reflected as an improvement in the subsequent criminal career, either a reduction in λ or a shortening in the duration of the career, as shown in Figure 4.

Alternatively, those who argue that prison is a "graduate school for crime" would anticipate a criminogenic effect on the career. That criminogenic effect might be represented as an

FIGURE 2
BASIC MODEL OF AN
INDIVIDUAL CRIMINAL CAREER

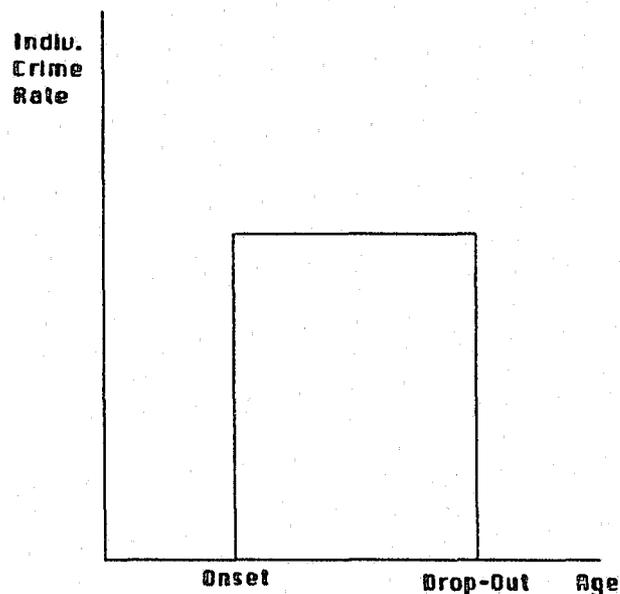
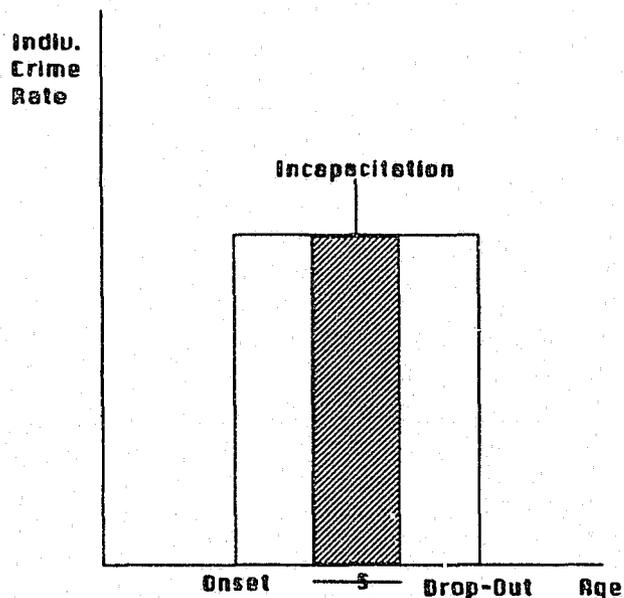
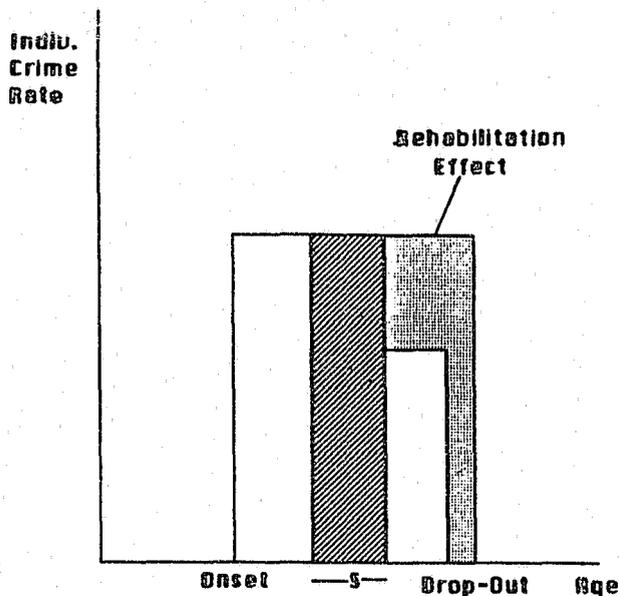


FIGURE 3
INCAPACITATION DURING A
CRIMINAL CAREER



extension in the duration of the career or an increase in the annual frequency, λ , as shown in Figure 5. This would diminish the crime-control effect of the imprisonment.

FIGURE 4
POST-RELEASE CONSEQUENCES ON
CRIMINAL CAREER PARAMETERS:
REHABILITATION



The incapacitative effect is also diminished if the individual is imprisoned for a sufficiently long time or sufficiently late in the career that the sentence S extends beyond the time when the criminal career would have terminated, a situation depicted in Figure 6. In that situation, the prison time after the end of the career is "wasted," at least from the viewpoint of incapacitation, although not necessarily so from the perspectives of retribution or deterrence. Thus, one has to consider the relationship between the residual criminal career and the duration of the sentence, S .

We can learn about criminal careers by observing two important stochastic processes: the arrest process and the crime process, and some symbols to distinguish their respective parameters are shown in Table 1. It is the crime process we ultimately care about, but most record data and much of the data used in research on criminal careers derives from the arrest process. The crime process can be observed through victims' reports to the police (which are reported in the annual Uniform Crime Reports or UCR) or to the National Crime Survey (NCS) victimization survey). These provide us with the annual crime rate, but with very little information about the individual offenders. The counterpart aggregate rate from the arrest process is the arrest rate per capita per year (A).

The individual crime-event rate λ has a counterpart individual arrest rate, μ , the arrest rate per active offender per year, which obviously must be lower than λ . The linking relationship connecting the arrest process and

FIGURE 5
POST-RELEASE CONSEQUENCES ON
CRIMINAL CAREER PARAMETERS:
CRIMINOGENESIS

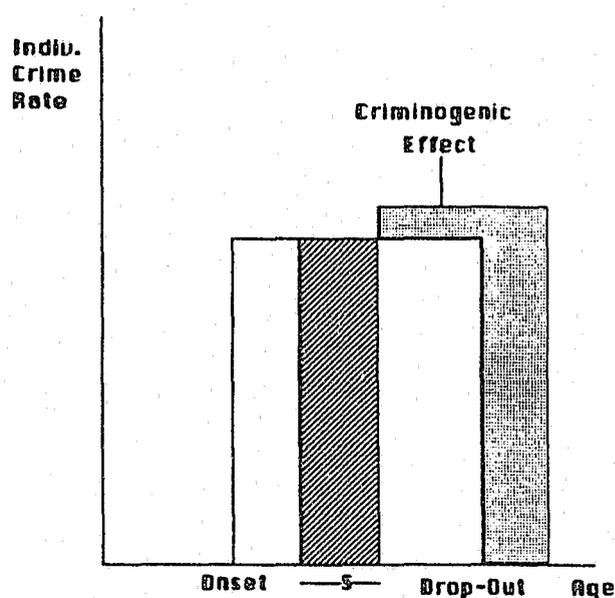
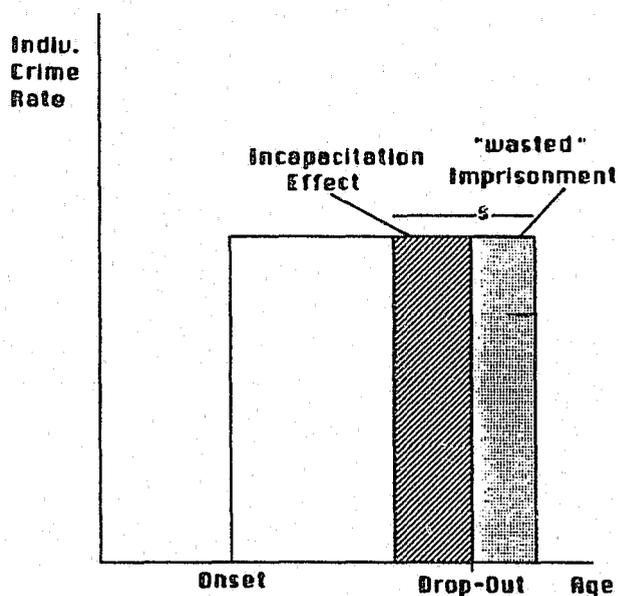


FIGURE 6
INCAPACITATION AFTER TERMINATION
OF A CRIMINAL CAREER



the crime process is the identity $\mu = \lambda q$, where q is the arrest probability per crime. Thus, for example, if $\lambda =$ ten crimes per year and q is .1 arrest per crime then one would anticipate seeing active offenders arrested at the rate of $\mu = 1$ arrest per year.

TABLE 1

	Crime Process	Arrest Process
Event Frequency	λ	μ
Aggregate Rates	C	A
Current Participation	d	b
Cumulative Participation	D	B
Basic Relationship	$C = \lambda d$	$A = \mu b$
Linking Relationship	$\lambda q = \mu$	

The issue of participation or prevalence is also relevant. Within the crime process, the fraction of the population doing crime is represented by "d", (mnemonically linked to "doing"), with a counterpart b (for "busted") for participation in the arrest process. The cumulatives (ever doing or ever arrested) are D and B, respectively. The basic identity in the crime process was indicated earlier as $C = \lambda d$, and the counterpart in the arrest process is $A = \mu b$. The crime and arrest processes are linked through the arrest probability, q, conditional on committing a crime so that μ , the arrest rate, is simply $\mu = \lambda q$.

There is obviously a relationship between the observation period over which one measures the doing of crime and the fraction who get busted, a longer observation period asymptotically increasing the probability of at least one arrest.

I would like to report some empirical results on some career parameters. Figure 7 (2) is an estimate of the fraction of males living in cities who ever get arrested for an FBI index crime (3) by age a. This is the function, B(a).

The prevalence numbers here are strikingly larger than most intuition suggests. There are also clear differences with race. For white males in a U.S. city, the chance of getting arrested for an index offense sometime in his life is about 15%, whereas for a black male it is over 50%. The quick rise occurs during the teenage years, when most first arrests occur.

Question: Is that self-reported?

Answer: No, this is based on official arrest data. We took Wolfgang's Philadelphia(4) data and extrapolated it to the demographic characteristics of the 55 largest U.S. cities. I should emphasize that those other cities had arrest rates that were in no sense significantly larger than Philadelphia.

One of the valuable sources of data on λ comes from self-reports by offenders--the only direct way to observe the crime process.

FIGURE 7. Ever Prevalence: Probability of Index Arrest by Age (a)

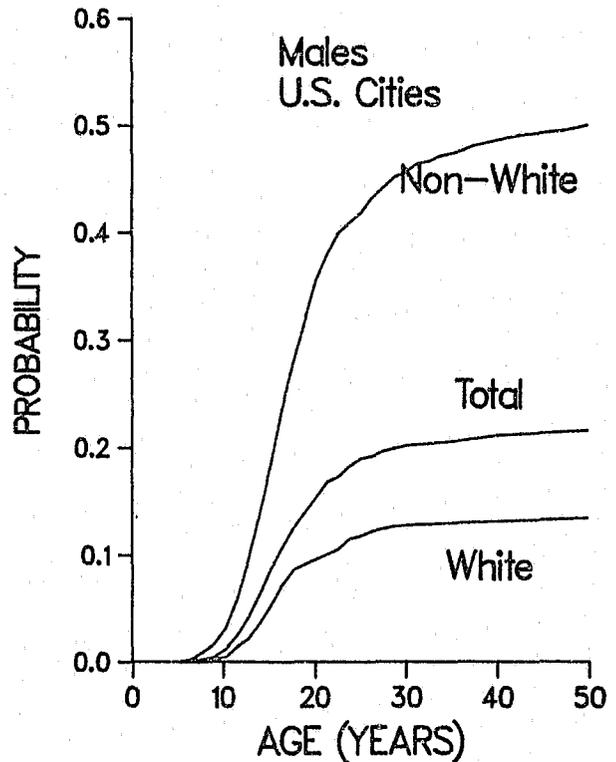
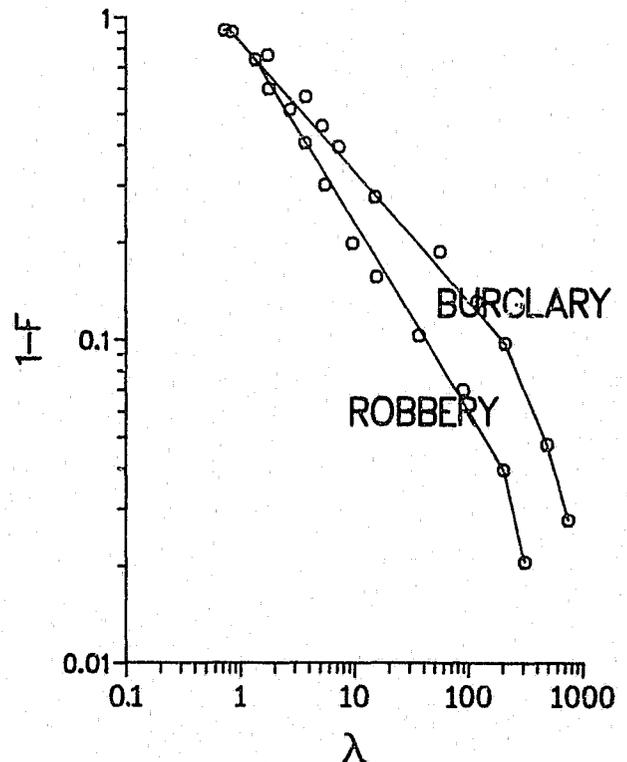


FIGURE 8. Distributions of Lambda in Rand Inmate Study



Rand(5) did a survey of prison inmates and tried to elicit their individual values of λ . Prisoners are an admittedly selected population, but if q is at all reasonably homogeneous in the population, the prison population should contain relatively high- λ folks.

The results of that survey were particularly intriguing because the distribution of λ they found was extremely skewed. I have plotted the distribution in Figure 8 on a log-log scale (with λ on the abscissa and $1-F(\lambda)$ on the ordinate) to capture the skewness. The skewness is reflected in the fact that the medians in this population were in the order of 5 crimes per year, but the 90th percentiles were in the order of 50 robberies per year and 200 burglaries per year.

The distributions show considerable bunching at the left and extremely long right-hand tails. One can raise questions about the reliability of those data, whether the deniers are contributing to the left-hand bunching and braggarts to the right-hand tail. There are artifactual questions about the degree to which the prisoners were caught during a short interval of spurting activity that was reported in the self-reports, and was then wrongly extrapolated to an annual basis. Putting all these questions to rest will obviously require further replication with more attention to these issues. My sense, however, is that the results are reasonably credible; there is a smoothness to the distribution in the middle. Also, the saturation effect shown in the right-hand tail could reflect constraints on the physical rate of committing crimes; that tail also suggests that here, as in many other phenomena, we know that there are relatively few people who do function at very high rates. It is clear that this is by no means the final story on estimates of the λ distribution, and there can be many corrections to it. It is, however, a very important first estimate of the λ distribution for at least one important sample.

Question: How long were they asked to recall? Their entire criminal career?

Answer: No, they were asked to report on the number of crimes they committed in a window of time that was between one and two years prior to the date on which they were arrested for the offense that led to their imprisonment. They certainly didn't try to integrate over the whole career. Rand tried a variety of steps to try to get an estimate of an annual rate.

The data force us to recognize that a large majority of offenders are bunched at the low end, and that is particularly important because prisoners are likely to be the worst offenders. Most of the attention has been paid to the right-hand tail, of these the high- λ people, who are the most important offenders. Also, the left-hand tail undoubtedly includes many deniers--people who simply deny what they did. That would tend to move the

median up somewhat. The means of the distributions, of course, are driven predominantly by the right-hand tail, and so the estimates of the means tend to be very high.

This partitioning of criminal-career parameters permits us to examine a number of covariates of career parameters. Figure 7, for example, showed some rather striking racial differences in cumulative participation rates. Those same large differences are reflected in the aggregate arrest rates, as shown in the first column of Table 2. Black per capita arrest rates are 4 to 10 times those of whites. It is important to examine how λ (the individual crime rate) or μ (the individual arrest rate) vary when one focuses on only the active offenders. The second column of Table 2 shows that the rates for blacks and whites who are active are much closer to each other.

TABLE 2

BLACK/WHITE ARREST RATIOS

	Aggregate Arrest Rate (R)	Arrest Rate for Actives
All Index	4.5	1.5
Property	4.0	1.6
Violent	5.2	0.97
Robbery	10.1	1.7

Thus, we see that participation is a very important filter. At any given time, within any population subgroup, a relatively small fraction of is currently active. The filter is a tighter one for whites, and so proportionately fewer whites become active offenders compared to blacks. The participation curve of Figure 7 shows that. But when you examine the subset that does penetrate the filter, the ones who are indeed active--black offenders and white offenders look much more similar. And, of course, the relevant decisions of the criminal justice system are focused predominantly on active offenders. The criminal justice system rarely gets to see members of the base population who are not active in crime. Rather, they only get to see the active subset. As a result, insights and judgements that are derived from aggregate rates can be very misleading when they get to focus on the actives, the ones about whom their decision making is most relevant.

I have not yet said very much about the "envelope" of the career, and that includes issues of career length, the initiation process, and the termination process. Initiation is predominantly a phenomenon that goes on in the teens. This was reflected in the rapid rise through the teen years in the participation curve (Figure 7), which rose close to its asymptote by age 20. One would also be inter-

ested in knowing more about the patterns of dropout and residual career length (which is reciprocally related), and in knowing how that varies with age and with other parameters.

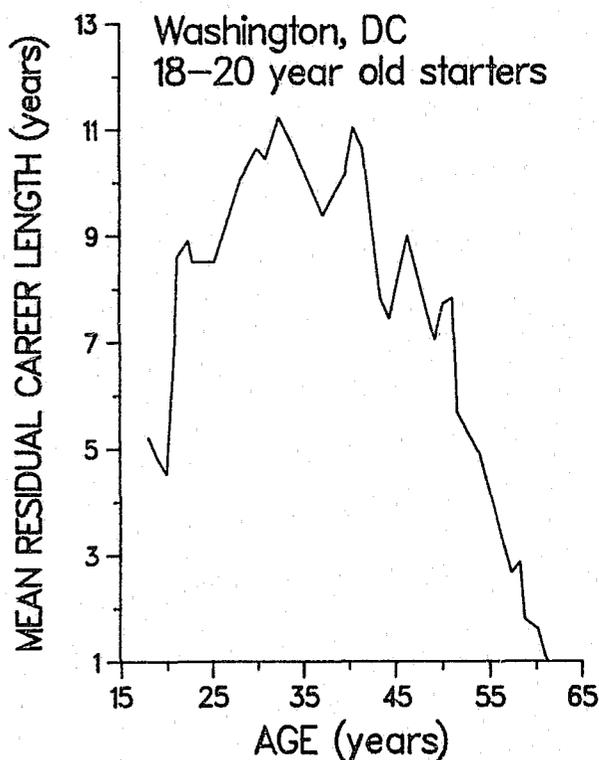
One of the ways that one could estimate career length is to go to official records and look at when people stop getting arrested. But, of course, the population observed has to be old enough so that a reasonable number of people will have stopped. Aside from the concern that that information will then be obsolete, one can never be sure when someone has stopped. A long interval since the last arrest may simply reflect only a temporary pause rather than a termination.

One can try another approach similar in construct to what one does in a life-table analysis. (6) The starting point here is the age distribution of arrestees, and then recognizing that if there are fewer 30-year-old arrestees than 25-year-old arrestees (as there are) a number of factors may account for the drop. It may be that μ is declining with age; if individual μ 's can be estimated by age, then one can calibrate for that effect.

It is also possible that the base-population cohort that is now 30 is smaller than the one that is now 25; obviously, population statistics can account for that. Finally, one is left with the explanation that many of the 25-year-old offenders terminated their careers by age 30. Adjustments for the other explanations can be used to isolate estimates of the dropout rate as a function of age. The information can then be used to develop estimates of the residual career length, the expected number of years left in the career at any given age.

These estimates were generated using data from Washington, D.C. The estimate of mean residual career length with age (Figure 9) is a particularly striking contrast to the aggregate arrest rate with age that was shown as Figure 1. In particular, mean residual career length is "increasing" from the start of the adult career until about age 30 and seems to reach a maximum at about age 30. This is consistent with the basic construct of initial break-in failures so familiar from reliability theory. The notion is that the "weak of heart" are increasingly being weeded out in the early years of their offending careers. The subset that is left after those early ones have dropped out are the more determined, the more persistent, and the ones most likely to stay. As one examines Figure 1, one notes that not many robbery or property offenders are left at age 30, but Figure 9 tells us that the ones who are left are the ones who are relatively more committed to crime and relatively more likely to continue their criminal careers. Thus, in contrast to the conventional wisdom of high termination rates at age 30 that derives from aggregate statistics, when one focuses on the actives we see that career length is at its maximum value throughout the 30's, and is relatively flat.

FIGURE 9. Mean Residual Career Length.



We then do notice rather rapid termination in the 40's. That may be associated with literal mortality, but mortality of normal populations at those ages (about 2 per 1000) is much lower than could account for the rapid termination observed. Even the higher mortality of criminal populations (say, parole releasees) is in the order of 10 per 1000, and the observed career termination rate is about 200 per 1000. The high termination rate may be associated with some kind of male menopause, a decaying of the skills and the motivation necessary to maintain the criminal career.

Question: Did you adjust for incarceration?

Answer: These estimates are not adjusted, but let me pursue that issue because it is a good one. If the age distribution of incarceration were close to the age distribution of arrests, then no adjustment is necessary because it is simply a minor scale difference. If the age distribution of incarceration is different--which it is, particularly for blacks--that requires some adjustment. But the adjustment is a relatively small one because the number of people in prison at any given time is relatively small. We have the data for the correction, but it is not reflected in Figure 9.

Question: Why mean distribution rather than median?

Answer: In dealing with λ , one is able to generate the full distribution, and so we can generate any statistic you want. Generating the

entire distribution of career length is rather more difficult, and so we have simply been calculating the means. Generating the median of the distribution would depend on the shape of the distribution and it might be useful, but I am not sure that we could get the full distribution comfortably. I am not comfortable with estimates of the tails because we are generating the estimates based on drop out rates as a function of age.

We are particularly interested in the determinants of the parameters. There is a long history of research on the "correlates of crime," or identifying the various elements of the set X in the relationship $C(X)$. When we write $C(X) = \lambda(X_1) d(X_2)$, however, the elements of the sets X_1 and X_2 could well be very different. Indeed, some of the earlier discussion called attention to some of the ways in which they are very different.

The criminal-career parameters we have been focusing on are clearly related to the measures that people have been dealing with for a long time. The most popular traditional measure for looking at criminal careers prospectively is the recidivism probability, the probability that someone will be rearrested in some exposure period. Of course, that measure depends on the arrest activity rate, μ . Someone could fail to be rearrested in an observation interval if he was simply lucky in the arrest process, or he could fail to be rearrested if the career terminated, at which point μ drops to zero. So, one can thus develop the relationship between this recidivism probability and the parameters μ and δ (the drop out rate) under reasonable distributional assumptions.

I want to make some observations about some of the problems of collecting data to estimate the various parameters discussed here. The ideal solution would be to recruit a random sample of offenders to keep daily logs of their criminal activity, and submit those logs to an appropriate reporting station reliably, honestly, and fully. The inadequacy of this ideal solution is what makes criminology a fascinating field of inquiry: it would be utterly dull if this solution worked. Instead, there are two realistic solutions. One is the arrest process, the one we have talked about predominantly. Here, the arrest process can be viewed as a sampling or a thinning of the underlying crime process that is of primary interest. And in order to know more about that, one has to learn about the sampling probability, q . To the extent that q is indeed homogeneous across the population, then we could estimate q from aggregate data and our job would be largely done. Unfortunately, q is not homogeneous across the population and so we have a difficult problem of finding covariates of q in order to make estimates of the underlying crime process from the data on the arrest process.

Fortunately, there is another way of observing the crime process, and that is through self-reports that involve asking people about the crimes they committed. This provides another window on the underlying crime process through the personal recall of the reporting individuals.

Each of these two sources of data clearly has important flaws. It is important to contrast them and to use both symbiotically, since neither can be viewed as definitive source. Together, hopefully, one might get a much better triangulation on the underlying process than either could provide alone. The self-report data, of course, is complete in that it is not merely a sample of crimes. It can be distorted, however by a non-response bias. If the refusal to provide self-reports is at all positively related to their rate, λ , then those who do report are a biased sample who commit crimes at a lower rate than the larger population. Also, people may forget, or they may have difficulty remembering whether a crime occurred within or before a particular observation period. This recall error can be compounded by misrepresentation at both ends, the braggarts at the high end and the deniers at the low end.

In contrast, self-report data are particularly rich because a wide variety of covariates can be linked to the self-reported information on offending. Many aspects of individual life histories and of individual attitudes can be important explanatory variables. In contrast, official records are much more limited in terms of the variables that they record. This richness of data does not come free, however, since the interviews that yield the data require an interviewer to elicit the responses, and so for any fixed budget, the sample size is limited.

In contrast, official records of the arrest process provide only a sample rather than the full array of events any individual did. The sampling bias is unknown and the fact that offenders have heterogeneous arrest vulnerability could be a problem. This could be compounded by recording bias; even if arrest vulnerability were homogeneous, the arresting police officer may decide not to arrest someone who is of similar ethnicity or whose family he knows, and to record the arrest of someone who is different.

In order to infer the characteristics of the underlying crime process from the arrest process, one has to make a variety of assumptions, as one must do with all inference. Some of those assumptions may be questionable.

In contrast to the richness of variables available in self-reports, where one can get individual juvenile experiences, school experiences, parental experiences, the variables as-

sociated with official record data are much more limited. On the other hand, cost considerations very much favor official records. Most individual arrest histories currently exist in machine-readable form.

A major policy use for criminal-career information is for consideration of incapacitation policy. A useful relationship was derived in a paper by Avi-Itzhak and Shinnar in 1973 (7). Their model starts with a number of simple assumptions about a Poisson crime process, exponential career length, and time served in prison, and general homogeneity of all the parameters of the crime and incarceration processes. They then derive estimates of the fraction of crimes averted through an incapacitation policy. This is reflected as the fraction of an individual's criminal career that is spent in prison:

Percent time incapacitated

$$= \frac{\lambda q_c QS [T / (T+S)]}{1 + \lambda q_c QS [T / (T+S)]}$$

where:

q_c = Probability of conviction given crime

Q = Probability of imprisonment given conviction

S = Expected time served

λ = Individual crime rate

T = Career length

The key parameters that enter include the probability of conviction given a crime, which aggregates everything involved in going from the crime event to the conviction event. In addition, they include the probability of imprisonment given conviction, the expected time served on a sentence; λ , the individual crime rate; and T , the career length.

This relationship provides a starting point for taking the criminal-career information on λ and T , and linking that to some very real policy considerations associated with the effect of policy variables, especially Q and S , on the incapacitation effect.

It should be clear that the base of empirical research as well as policy usefulness is still at a very primitive level. Even in the last few years, however, there have been some very significant developments both in estimating the parameters, in generating insights about their determinants, and in providing an environment for linking information about individual criminal careers to appropriate policy choices. The efforts have only recently begun, and very much more remains.

John Lehoczky

In the time remaining, I want to look at a few more specific and lower level issues in criminal career modeling. To try to give an overview of what I intend to say, I would like to focus on why we want to develop models of criminal careers and what aspects should be included in such models. If we refer to Figure 9 giving the Rand self-report survey data, we see tremendous heterogeneity of offender behavior. It is useful to conceive of this using a statistical concept arising in the analysis of variance: variation within and variation between. We see in Figure 9 a very large variation between individuals. Given any particular individual we want to understand the variation for single career given all the parameters that were defined earlier by Professor Blumstein: the crime rates, arrest probabilities, dropout times and so on. Each individual will have a stochastic process realization given these parameters, and this can be thought of as the variation within a single career. In addition, there is a great amount of variability from individual to individual which cannot be explained by a stochastic process alone. We want to include both sources of variability in our models in order to develop good policies for reducing crime rates and to predict the consequences of any of these policies, so that we have some sense of how they will perform. Both the variability within and the variability between must be taken into account in any sensible model of criminal careers.

There is a second topic that I want to address; that is, to try to identify, understand, and then possibly remove biases that are inherent in many of the criminal justice data sets that we rely on to estimate criminal career parameters. Some biases arise from the way a dataset is collected. Others can be accentuated by the variability among individuals. The approach that recognizes this variability will lead to methods which aid in the task of reducing biases to obtain better statistical inferences.

If we go back in the literature of basic models of criminal careers. We will find starting with the Avi-Itzhak and Shinnar paper that Al alluded to earlier, that most of these papers use very simple kinds of probabilistic methods. This is done so that one can get a tractable kind of model for deriving behavioral results. The models generally assume that an individual commits crimes at some rate λ , so the crime process is usually treated as a Poisson process or perhaps a slightly more general version, a renewal process. There isn't any accommodation of age dependency even though it is generally accepted (see Figures 1 and 9) that there is a serious age dependence in the crime commission phenomenon. With re-

spect to the arrest process, it is conceived of as being just a coin flip. An offender commits a crime and then an imaginary coin is flipped to determine whether or not that individual will be arrested. The arrest process becomes what is called a thinned version of the crime process. To the extent to which offender dropout is considered at all in the early models, it is modeled as an exogenous event. That is, an offender starts a criminal career, but in his pocket is a stop watch. When an appropriate amount of time has elapsed, an alarm rings and the career is over. The dropout time is independent of any aspect of the career no matter how many times the offender is incarcerated or for how long. Finally, the most important limitation is the assumption that all individuals are characterized by identical parameters. We may now know the parameters values and so we can use data to estimate them: however, it is assumed that individuals are all homogeneous in all of these aspects: crime rates, arrests, probabilities, conviction rates, and dropout rates. Looking at criminal justice data sources, it becomes fairly clear that there are at least 4 important aspects which should be included in these models but are not. First, crime rates do vary greatly with age. Recalling Figure 1, and especially burglary and robbery data, we know that there is a profound age dependency. Also from the self-report data we know that the offender population is extremely heterogeneous in crime rates and in crime types. Third, we know there are many different kinds of offenders, including offenders in property crimes, offenders in violent crimes, specialists, and generalists. Also crime commission can be a bursty activity. There can be periods of time in which an offender is extremely active and then quiescent for other periods of time. None of the early models take this phenomenon into account. Finally, there is some data indicating that dropout really isn't governed by the stop watch, alarm clock model. In fact, one thing that might well be true is that an offender begins with many different kinds of crime types but systematically stops doing certain kinds of crimes. He may be active in many different crime types early in his career but slowly stops doing certain crime types. The effect of such behavior would be to create an age dependency. When an offender is young he commits many different kinds of crime types and so would tend to commit many crimes of various sorts. As the possible types are eliminated, the offender's crime rate will diminish. Models generally don't address this point.

Now, to be more specific concerning examples of biases, in criminal justice data sets there are a variety of different data sources that one might use to try to estimate various aspects of criminal careers. One that has been used at Carnegie-Mellon is window arrest data from Detroit. One observes a window of time and within that window of time gathers data on every individual who is arrested. For each such individuals, we obtain their entire criminal record up to that time. Then you use all

those records that you garnered to try to estimate criminal career parameters, crime rates or whatever the parameters of interest might be that have been mentioned here. This is a really rich data source but problems arise. One issue is that the data that I just described is gathered conditional on an event. In this case, the event is being arrested in whatever the particular window of time happens to be. That is, an individual is included in the data set conditional on being arrested in the time window. Consequently, we must realize that we are most likely to be catching high rate offenders, because high rate offenders are more likely than low rate offenders to be captured in the window. So, if you look at those records, you are not going to be looking at a random sample of offenders in a population but rather a special kind of filtered random sample which tends to have high rate offenders. I say "tends to have" because low rate offenders can be caught too, but on balance the high rate offenders have a much greater chance of being captured. Also, if there is a bursty nature to crime commission, then you are likely to be catching somebody in a burst as opposed to those individuals who are in a more quiescent time and are not likely to be caught inside the window. Consequently, we need to observe that the data is collected conditional on an event and the probability of the event involves the parameter. It is difficult, in one step, to validly conclude anything about the offender population. One can draw conclusions about the population that was captured, but that is less interesting than conclusion relating to the offender population in general. When one thinks about crime control policies, one is trying to apply them to the offenders at large rather than the ones that happen to fall into any particular window.

FIGURE 10: WINDOW ARREST BIAS

	0	1	2	3	4	5
1 $\lambda=0.25$		x				
2 $\lambda=0.58$		xx				x
3 $\lambda=0.54$						x
4 $\lambda=0.66$		x	x		x	
5 $\lambda=1.22$	x		x	xx	x	
6 $\lambda=0.09$						
7 $\lambda=2.00$		x	xx		x	
8 $\lambda=0.22$			x			
9 $\lambda=0.31$						x
10 $\lambda=3.19$	x	x	xxxxx	x	xx	xx

Cases 4, 5, 7, and 10 have window arrests, others do not. Case 4 had his 3rd arrest at time 3.51, 5 had his 6th arrest at time 3.66, 7 had his 4th arrest at time 3.12 and 10 had his 14th arrest at time 3.77.

To further illustrate the problem with window arrest data, I performed a very simple simulation experiment. Imagine 10 individual offenders, each with a randomly generated crime rate parameter. The rate parameters were generated according to an exponential distribution (see Figure 10). The ten parameter values were 0.25, 0.58, 0.54, 0.66, 1.22, 0.09, 2.00, 0.22, 0.31 and 3.19. Each offender is now endowed with a crime rate parameter λ . Now we turn each individual loose to commit crimes. Each commits crimes according to a Poisson process with his individual rate parameter. Now imagine that there is a window of time, say [3,4] and that any individual committing a crime in that window is caught.

Referring to Figure 10, we see that individuals 1,2,3,6,8 and 9 have no arrest in the time window [3,4], while individuals 4,5,7 and 10 do. So we actually catch 4 out of the 10, and then we get to look at the four records. We consider each of the four records and try to make some inferences about crime commission rates. It turns out that the fourth individual was caught on his third arrest at time about 3 1/2. Number 5 was caught on his sixth arrest. We caught number 7 on his fourth arrest, and we caught number 10 on his fourteenth arrest.

Notice first that the four individuals who were apprehended were the four with the largest values of λ ; 3.19, 2.00, 1.22 and 0.66. The average λ value is 1.79, substantially larger than the population average of 1.00. In reality, one would not be given the values of λ but would be forced to estimate them from the historical data. A crude, but standard, estimate for each individual λ might be total number of arrests divided by the time up to the window arrest. This would yield estimates of 0.86, 1.64, 1.28 and 3.71 for the four individuals. Assuming a model with a single λ value for each individual, we would obtain a combined estimate of 1.92. This would be nearly twice as high as the population value of 1.00. This analysis is, of course, incorrect because it takes no account of the heterogeneity of the offender population and ignores the fact that the data is gathered conditional on having a window arrest. This makes it far more likely that high rate offenders will be caught, thus biasing the results. Fortunately, there are very simple procedures available to correct for this. One needs to write the likelihood function conditional on having an arrest in the interval [3,4]. This correction will bring the estimates back into line.

There is a second issue which is not brought out in this example where individual crime processes were assumed to be Poisson processes. Crime (and arrest) processes tend to exhibit bursty activity; that is, crimes are committed in clusters. If one gathers data for study from window arrests, then it is likely that this will be part of a burst of activity. Consequently, simple estimates of crime rates will be inflated by the presence of a burst.

Again, this difficulty can be corrected by developing a better stochastic process model for this crime commission process.

A second kind of bias I am currently investigating arises with prison survey data. If one considers surveys of prisoners, (e.g., Survey of Inmates of Correctional Institutions. (1974 or 1979) one is given a snapshot taken during a particular one month or two month time period of a wide variety of correctional institutions. If we consider the prisoners involved in the survey, we know immediately that they are in no way a random sample from the offender population. They are very special in that they are in prison. In order to get into prison you have to commit crimes and enough of them in order to have the privilege of going to prison. So we know that in general these individuals have long arrest records that finally get them to prison or they have committed the types of crimes that gets one to prison. Consequently, there are very obvious kinds of biases in those data sets, and one must be very cautious in terms of extrapolating back to any estimates of the offender population or to any type of calculations one might make about the efficacy of particular crime control policies. Let us look at one aspect of this problem to show a case where modeling can be very helpful. There is an obvious length-biasing problem. Suppose we are interested in estimating sentence lengths from these data. Imagine that rather than sampling, we take a census of the prisoners and study their sentence lengths. We will obtain a very biased view of sentence lengths unless we correct for the length biasing phenomenon. The problem is that those with long sentences could be overrepresented in the census, therefore those with long records and more violent crimes will be overrepresented because they have longer sentences and more violent crime patterns. I say "could" because some survey designers, such as the Rand self-report survey, are aware of the length-biasing problems and have corrected for this kind of phenomenon. It is useful to note that this length-biasing problem would not be present if we had prison intake data; however, surveys may not be carried out in such a fashion.

Let us now consider a simple simulation experiment. Consider a prison with ten cells in it. As soon as a cell becomes empty because a prisoner has finished his sentence, a new prisoner appears and occupies the cell. Suppose that there are two types of offenders. There are offenders who are sent for short periods of time, in this case a time of one unit, so one time unit after they go into jail, they are released. There is another type of offender, a high-rate offender. These offenders are sentenced to five time unit sentence lengths. Assume that at time of intake, there is a .5 probability of the offender being high (or low). Thus on average half of those who come in are high and half who come in are low, and they stay for whatever the high time is (5 units) or for whatever the low time is (1 unit). Figure 11 gives simulated data.

FIGURE 11: PRISON SURVEY BIAS

	0	2	4	6	8	10	12	14	16	18	20
1	L	L	L	L	H	L	H	H			
2		H		H		H	L	H			
3	L	L	H		H		H		H		
4	L	L	H		H	L	H	L	L		
5		H		H		H	L	L	L	H	
6		H		H	L	H	L	H			
7	L	H	L	L	H		H	L	L		
8	L	H		H	L	L	H			H	
9	L	H		H	L	H		H			
10		H		H	L	L	L	L	L	H	

Two types of offenders:

Low (probability = .5) Length = 1
 High (probability = .5) Length = 5

Consider cell #1. The first four occupants are low rate. Between times 4 and 9 it is occupied by a high rate offender. This is followed in turn by a low rate offender and then two highs. Each of the other 9 cells are generated similarly. Each choice of H or L is made randomly with probability .5.

If we look at the composition of the cells at time 0.5, we see that 6 are low while 4 are high. This is in accord with the assumption that the initial times are chosen at random with equal likelihood. Now look later in time, say at time 19.5. Here only two of the cells (#4 and #7) are occupied by low rate offenders. The other 8 are occupied by high rate offenders. Even though the intake rate is .5, we end up with an 80% high offender occupancy rate. Is this 80% rate some unlikely chance event? Actually, this is a well-known phenomena in renewal theory which is called length-biased sampling. Highs have five times the chance of being in a cell than lows have, so actually we would expect 5 out of 6. If I had used twelve cells rather than 10, then on average 10 would be filled with highs and 2 with lows. If you consider any time well away from the start time of zero, you will find that most everybody in the census is high. Consequently, one gets a very biased view from a census taken over a short time interval in a prison. If one is working with this kind of data set, then one must be alert to this phenomenon and correct for it.

I have been working with Professor Blumstein's Panel on Criminal Careers to create probability models that are richer than the ones that I have described earlier today. The models are richer in the sense that they include heterogeneous behavior, they allow for quiescent

periods and spurt behavior, and they introduce age dependencies. I have been interested in developing tractable models that include many of these extra effects. This was done in a paper which is included as a part of the Panel report and will appear in the Fall of 1986. I will not discuss the detailed model here, but will instead discuss hierarchical modeling.

Hierarchical models are used to allow for heterogeneity in criminal behavior. We assume that all of the parameters that we alluded to earlier such as lambda, the dropout times, arrest rates, etc., are themselves all random variables. That is, there is an offender population and we imagine that individuals are endowed with particular parameter values. There can be a big spread of parameter values across the population. Once endowed, the offenders pursue criminal careers using their own rates. We view these parameters themselves as being random variables, and the distribution of those parameters tells us something about the heterogeneity of the offender population (that is the variation between) offenders. If we can get a characterization of that population which I will call the superpopulation, because it is what is characterizing the differences among offenders, then we have gone a long way towards being able to analyze the effect of various policy changes. We would then be able to examine how policies will impact the entire offender population. The appropriate estimation technique is called the empirical Bayes approach. This approach leads us to use all records for all offenders to sharpen estimates for individuals. Usually one thinks about estimating an individual's lambda by looking only at that individual's record. What else could be relevant? Other offenders are committing crimes at other rates. If one uses the hierarchical modeling approach, then the records of other offenders become relevant and lead us to adjust individual estimates. They are relevant because they can be used to estimate the superpopulation distribution which, in turn, helps to estimate each individual's parameters. This can be controversial, because there is a moral-ethical issue of letting other criminal records influence the estimates of and treatment of a single offender. The empirical Bayes methodology is a very active research area in statistics these days. There will be numerous talks on it during the next four days. It seems to hold substantial promise for criminal justice system parameter estimation.

I carried out a small simulation experiment to illustrate the empirical Bayes method. Consider 10 individuals and draw a value for lambda for each. In this case, an exponential distribution was used. The 10 values were, respectively: .93, .07, .48, .10, .30, .17, 2.01, 2.17, .11, and 1.63. Next, each individual commits crimes according to a Poisson process with his chosen lambda. Suppose we have five interevent times for each individual. These are listed in Figure 12 and displayed in Figure 13.

FIGURE 12

Example of Empirical
Bayes Analysis

10 individuals, each has 5 observed inter-arrest times

1.	1.0969,	2.8249,	0.3512,	0.2585,	0.8455
2.	13.7771,	1.7669,	12.9052,	4.3913,	6.1609
3.	0.6495,	3.0787,	1.4430,	0.6609,	7.7858
4.	3.2332,	2.2115,	1.5468,	5.8593,	5.9848
5.	0.6918,	0.3450,	1.8421,	1.7785,	2.1152
6.	8.4208,	0.0526,	4.2800,	6.3074,	0.4596
7.	0.0340,	0.4107,	0.3995,	0.2592,	0.1365
8.	0.1590,	0.3802,	0.2753,	0.0959,	0.3616
9.	7.3398,	0.4768,	17.3890,	2.9763,	11.3156
10.	4.2388,	0.2154,	0.0904,	0.7811,	1.1245

The goal is to estimate the 10 individual lambda values and the mean of the lambda values for the entire population. The results are shown in Figure 13.

FIGURE 13

Estimates

	\bar{X}	$1/\bar{X}$	Bayes	True λ
1.	1.0754	0.9299	0.9409	0.9314
2.	7.8003	0.1282	0.1500	0.0704
3.	2.7236	0.3672	0.4105	0.4845
4.	3.7671	0.2655	0.3025	0.1020
5.	1.3545	0.7383	0.7720	0.2957
6.	3.9041	0.2561	0.2924	0.1720
7.	0.2480	4.0326	2.6787	2.0099
8.	0.2544	3.9308	2.6408	2.1716
9.	7.8995	0.1266	0.1482	0.1076
10.	0.7752	1.2900	1.2506	1.6296

$$\sum (\lambda_i - 1/\bar{X}_i)^2 = 7.5486$$

$$\sum (\lambda_i - \hat{\lambda}_i(\text{Bayes}))^2 = 1.1218$$

We consider two estimation approaches. First, we might use maximum likelihood methods. Each individual lambda would be estimated using the reciprocal of the average interevent time. The results are presented in the 1/X column of Figure 13. The total squared error for these estimates is 7.55.

The second approach is to utilize empirical Bayes methods. We first use all 50 data points to estimate the parameter of the exponential distribution. This estimate is 1.067, remarkably close to the true value of 1.00. Next, we assume that each λ_i , $1 < i < 10$ is generated from an exponential distribution with parameter 1.067. We take the five observations, compute the posterior distribution of λ_i and then estimate λ_i . The 10 individual estimates are given in the Bayes column of Figure 14. We note that the squared error is dramatically reduced from 7.55 to 1.12.

The empirical Bayes approach results in "shrinkage" estimators. The maximum likelihood estimators will be moved toward 1.067. For example, individuals 1,2,3,4,5,6 and 9 have 1/X values less than 1.067 while the corresponding Bayes estimate is larger (but less than 1.067). Each of these values is moved up toward 1.067. This leads to better estimates for individual #3 but slightly worse for the others. Conversely, the 1/X values for individuals 7,8 and 10 are above 1.067, and the corresponding Bayes estimates are substantially reduced. This leads to much better results for individuals 7 and 8. Overall, the total squared error is dramatically reduced, largely because of individuals 7 and 8.

We see that the empirical Bayes approach can lead to much better overall results. There are, of course, obvious ethical considerations. There are individuals who have averages which are below the mean. The empirical Bayes method will tend to elevate their values artificially. Individuals with high rates will have their estimates reduced. Nevertheless, this simple example shows how population variability can be incorporated into criminal justice models and the way in which such models can be properly estimated.

Notes

1. See, for example, Sheldon and Eleanor T. Glueck (1930); 500 Criminal Careers. New York: Alfred A. Knopf.
2. The figure is taken from Alfred Blumstein and Elizabeth Graddy, "Prevalence and Recidivism in Index Arrests: A Feedback Model" Law and Society Review, vol. 16 (1982), pp. 265-290.
3. The FBI's index crimes are homicide, forcible rape, aggravated assault, robbery, burglary, larceny, and auto theft.
4. The data were based on Philadelphia police contacts with boys born in 1945 who were in Philadelphia from 1955 (at age 10) through 1963 (at age 18), reported in Mawin Wolfgang, Robert Figlio, and Thorsten Sellin (1972); Delinquency in a Birth Cohort. Chicago, IL: University of Chicago Press.
5. The results of the survey were reported in Jan M. Chaiken and Marcia Chaiken, "Varieties of Criminal Behavior" (1982). Rand Report No. R-2814-NIJ. Rand Corporation.

6. This is the approach used by Alfred Blumstein and Jacqueline Cohen, with Paul Hsieh (1982), "The Duration of Adult Criminal Careers." Carnegie Mellon University.
7. Benjamin Avi-Itzhak and Reuel Shinnar. "Quantitative Models in Crime Control" Journal of Criminal Justice, vol. 1 (1973), pp. 185-217.

References

- Avi-Itzhak, B. and Shinnar, R. Quantitative models in crime control. Journal of Criminal Justice, 1, 185-217, 1973.
- Blumstein, A. Cohen, J., and Nagin D. (eds.) Deterrence and Incapacitation: Estimating the Effects of Sanctions on Crime Rates. National Research Council, National Academy of Sciences. Washington, D.C.: National Academy Press. 1978.
- Blumstein, A., Cohen, J., and Hsieh, P. The Duration of Adult Criminal Careers. Draft report submitted to National Institute of Justice, August 1982. Pittsburgh: Carnegie-Mellon University, School of Urban and Public Affairs. 1982.
- Blumstein, A., and Cohen J. Estimation of individual crime rates from arrest records. Journal of Criminal Law and Criminology 70:561-85. 1979.
- Blumstein, A., and Graddy, E. Prevalence and recidivism in index arrests: a feedback model. Law and Society Review 16:265-290. 1982.
- Blumstein, A., Cohen, J., Visher, C., and Roth, J. Criminal Careers and Career Criminals. Report of the NAS Panel on Criminal Careers; NRC/NAS, Washington, D.C. 1985.
- Chaiken, J.M., and Chaiken, M. Varieties of Criminal Behavior. Rand Report R-2814-NIJ. Santa Monica, California: Rand Corporation. 1982.
- Chaiken, J.M., and Rolph, J.E. Selective incapacitation strategies based on estimated crime rates. Operations Research, 28, 1259-1274. 1980.
- Cohen, J. The incapacitative effect of imprisonment: a critical review of the literature. Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions of Crime Rates, edited by A. Blumstein, J. Cohen, and D. Nagin, National Research Council, National Academy of Sciences. Washington, D.C.: National Academy Press. 1978.
- Deely, J.J., and Lindley, D. V. Bayes empirical Bayes. JASA, 76, 833-841. 1981.
- Elliott, D.S., and Ageton, S.A. Reconciling race and class differences in self-reported and official estimates of delinquency. American Sociological Review 45:95-100. 1980.
- Elliott, D., Ageton, S., Huiringa, D., Knowles, B., and Canter, R. The Prevalence and Incidence of Delinquent Behavior: 1976-1980. National Youth Survey, Report No. 26. Behavioral Research Institute, Boulder, Colorado. 1983.
- Farrington, D.P. The prevalence of convictions. British Journal of Criminology 2(2):173-175. 1981.
- Glueck, S., and Glueck, S. 500 Criminal Careers. New York: Alfred A. Knopf. 1930.
- Greenwood, P., and Abrahamse, A. Selective Incapacitation. Rand Report R-2815-NIJ. Santa Monica, California: Rand Corporation. 1982.
- Lehoczky, J. P. Random parameter stochastic process models of criminal careers. Department of Statistics Technical Report No. 343, Carnegie-Mellon University. 1985.
- Morris, C. Parametric empirical Bayes inference: theory and applications. JASA, 78, 47-65. 1983.
- Peterson, M.A., Braiker, H.B., with Polick, S.M. Who commits crimes: A survey of prison inmates. Cambridge, Oelgeschlager, Gunn, and Hain. 1981.
- Rolph, J.E., Chaiken, J.M., and Houchens, R.E. Methods for estimating crime rates of individuals. Rand Corporation Report R-2730-NIJ. 1981.
- Wolfgang, M.E., Figlio, R.M., and Sellin, T. Delinquency in a Birth Cohort. Chicago: University of Chicago Press. 1972.

Recent studies of race and victim effects in capital sentencing

George G. Woodworth
University of Iowa

Introduction

To begin this session on race of victim research, I will briefly present some of the quantitative results on the race of victim issue that have emerged from recent studies of capital charging and sentencing. One significant question addressed by these studies is whether homicide cases which are identical in all legally relevant ways but differ as to the race of the victim receive equal sentences.

I suggest that there are two approaches to the study of this question which although not in conflict, do differ in emphasis. The 'econometric' approach involves building a model of the capital charging and sentencing process while the 'epidemiological' approach involves case-matching.

In the model-building, econometric approach one attempts to construct a statistical model of the capital charging and sentencing system. The model would include legally relevant variables (such as contemporaneous felony), legally neutral variables (such as urban or rural location) and variables which may be legally impermissible as a basis for sentencing (such as race of victim). After completing the usual model building steps of checking model specification by residual analysis and other diagnostics, one examines the statistical significance of the racial variables in the model.

The case-matching, epidemiological approach involves forming subgroups of matching cases (all armed robberies, for example). Within each subgroup one then compares the death sentencing rates for white and black victim cases.

There is an important insight to be gained from a comparison of the two methodologies: to test if a variable such as race of victim influences the sentencing outcome, it is not necessary to produce a model which explains everything. What is instead required is that within each group of matched cases all relevant background variables have similar distributions in white and black victim cases. Rosenbaum and Rubin refer to this requirement as subclassification on the propensity score (1). In particular, mediocre predictive power is not by itself an indication of a problem with the analysis. Michael Finkelstein made the same point in a somewhat different way (2).

Modelling and case-matching approaches differ only in emphasis in my opinion, since a model could be used to match cases (by grouping together cases with similar predicted outcomes apart from racial influences). The results which I will shortly describe are based on both modelling and case-matching.

In the first study I will discuss, Gross and Mauro (3) analyzed FBI Supplementary Homicide Reports (SHR's) augmented with information from other sources. They separately analyzed data from eight states using case-matching and logistic regression modelling. The variables available to Gross and Mauro were location of the crime, presence or absence of a contemporaneous felony, the number of victims of the crime, whether a gun was involved, the relationship between victim and perpetrator and the races of the victim and perpetrator.

Table 1 (Gross and Mauro Table A-1) matches cases with respect to race of victim, number of victims killed by the perpetrator, race of defendant, contemporaneous felony, victim-perpetrator relationship, and use of a gun. There are many possible combinations of these variables each defining a group of cases, however not all combinations were observed. For example, in Georgia there were 45 nonempty groups of matched cases which could be formed into 19 matched pairs consisting of a white victim group paired with the corresponding black victim group. In seventeen of these matched pairs, the death sentencing rate was non-zero in the white victim group or the black victim group or both. If black and white victim cases were treated equally, one would expect the death sentencing rate to be greater in the white victim groups about as often as in the black victim groups. However, in 15 of the 17 matched pairs in which death sentences occurred, the rate was higher in the white victim group than in the matching black victim group, a disparity which is statistically significant at the .002 level.

Table 2 (Gross and Mauro Tables 24 and 32) illustrate a modelling approach. In this case logistic regression models of the death sentencing rate show statistically significant race of victim effects in several states. The logistic regression coefficient contrasts a perpetrator's odds of receiving a death sentence in a white victim case with the odds of receiving a death sentence in an identical black victim case. For example, in Georgia, killers of white victims are estimated to have seven times the odds of receiving death sentences as killers of black victims under similar circumstances.

TABLE 1

CELL-BY-CELL ANALYSIS OF DEATH SENTENCING BY RACE OF VICTIM

	GA	FL	IL	OK	MS	NC	VA	AR
1. Number of Non-empty Groups	45	54	58	38	38	42	46	41
2. Number of Matched Pairs of Nonempty Groups	19	24	27	14	15	17	20	17
3. Proportion of Cases in Matched Pairs	$\frac{2110}{2118}$	$\frac{3308}{3486}$	$\frac{2984}{3080}$	$\frac{766}{898}$	$\frac{799}{852}$	$\frac{1758}{1871}$	$\frac{1374}{1389}$	$\frac{775}{798}$
4. Number of Groups in Which Death Sentences Occur	36	42	34	31	15	23	19	16
5. Number of Matched Pairs in Which Death Sentences Occur	17	18	16	12	6	10	9	7
6. Proportion of Death Sentences Included in Matched Pairs	$\frac{76}{79}$	$\frac{107}{128}$	$\frac{39}{45}$	$\frac{32}{43}$	$\frac{18}{22}$	$\frac{20}{27}$	$\frac{18}{19}$	$\frac{12}{15}$
7. Proportion of Matched Pairs With White Victim DP Rate > Black Victim DP Rate	$\frac{15}{17}$	$\frac{15}{18}$	$\frac{13}{16}$	$\frac{11}{12}$	$\frac{5}{6}$	$\frac{6}{10}$	$\frac{8}{9}$	$\frac{5}{7}$
	.002	.005	.012	.004	.109	.055	.020	.226

Baldus, Pulaski and Woodworth (4, 5) examined a smaller group of homicides in one state, Georgia, but were able to control for a large number of background variables. Two sets of data were collected, the first study (Procedural Reform) covered somewhat more than 600 offenders convicted of murder at trial. The second, more extensive study (Charging and Sentencing) covered offenders convicted of murder or voluntary manslaughter. Information was obtained from records of the Georgia Board of Pardons and Paroles, the Department of Offender Rehabilitation, vital statistics, appellate and Supreme Court opinions and briefs and other sources. Preliminary analyses of these data were presented in August, 1983 in testimony in *McCleskey v. Kemp* which was subsequently argued before the United States Supreme Court on October 15, 1987.

A variety of linear and logistic modelling and case-matching strategies have been used in analyzing these data and are reported elsewhere. Two analyses of the charging and sentencing, data which were presented in the 1983 testimony are shown here. Figure 1 illustrates a linear regression model involving 39 explanatory background variables. The dependent variable was coded 1 if a death sentence was imposed and zero otherwise. With this type of dependent variable, a linear regression equation is a model of the probability of a death sentence as a function of the explanatory variables. The model is technically 'linear' in the sense of being a weighted linear combination of a fixed set of functions of explanatory variables. In this case the terms in the model are: level of aggravation (a composite of the 39 background variables) and its square, race of victim

TABLE 2

BEST LOGISTIC REGRESSION MODELS, SUMMARY STATISTICS

	GA	FL	IL	OK	NC	MS	VA	AR
Victim's Race: Logit Coefficient	1.97***	1.56***	1.38***	1.46*	1.47*	1.70**	0.88	1.27
Victim's Race: Multiplier of Odds of Death Sentence	7.2	4.8	4.0	4.31	4.35	5.47	2.41	3.56

*** Effect significant at or beyond the .001 level.

** Effect significant at the .01 level.

* Effect significant at the .05 level.

and defendant, and linear and quadratic interactions between race of victim and level of aggravation. Higher order terms involving race of defendant were not statistically significant. The curves in Figure 1 are 95 percent confidence limits for the mean probability of a death sentence for black defendant cases with black victims or white victims at increasing levels of aggravation. The key feature of this model, which recurs in a variety of other analyses, is that the race of victim disparity is concentrated in moderately aggravated, 'midrange' cases, where the death sentencing rates for white victim cases are over 20 percentage points higher than for similarly aggravated black victim cases.

Table 3 shows a case-matching analysis of defendants indicted for murder. Defendants have been grouped into eight increasingly aggravated groups. Rates at which prosecutors seek a death sentence and juries impose a death sentence are shown for white and black victim cases. Again, white victim rates exceed black victim rates, particularly in moderately aggravated 'midrange' cases.

FIGURE 1. Black Defendant Model, Georgia Charging and Sentencing Study

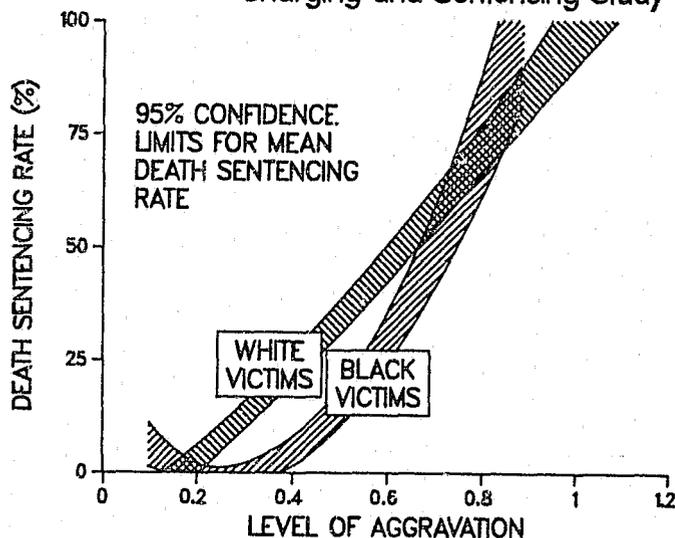


TABLE 3

RACE OF VICTIM DISPARITIES IN DEATH SENTENCING RATES AMONG DEFENDANTS INDICTED FOR MURDER CONTROLLING FOR THE PREDICTED LIKELIHOOD OF A DEATH SENTENCE AND THE RACE OF THE VICTIM

Predicted Chance of a Death Sentence l (least) to 8 (highest)	Average Actual Sentencing Rate For the Cases at Each Level	Death Sentencing Rates for Black Defends. Involving		Arithmetic Difference In Race of the Victim Rates	Death Sentencing Rates for White Defendant Involving		Arithmetic Difference in Race of the Victim Rates
		White Victims	Black Victims		White Victim	Black Victim	
1	.0 (0/33)	.0 (0/9)	.0 (0/19)	.0	.0 (0/5)	-	-
2	.0 (0/55)	.0 (0/8)	.0 (0/27)	.0	.0 (0/19)	.0 (0/1)	.0
3	.08 (6/76)	.30 (3/10)	.11 (2/18)	.19	.03 (1/39)	.0 (0/9)	.03
4	.07 (4/57)	.23 (3/13)	.0 (0/15)	.23	.04 (1/29)	-	-
5	.27 (15/58)	.35 (9/26)	.17 (2/22)	.18	.20 (4/20)	-	-
6	.17 (11/64)	.38 (3/8)	.05 (1/20)	.33	.16 (5/32)	.50 (2/4)	-.34
7	.41 (29/71)	.64 (9/14)	.39 (5/13)	.25	.39 (15/39)	.0 (0/5)	.39
8	.88 (51/58)	.91 (20/22)	.75 (6/8)	.16	.89 (25/28)	-	-

FIGURE 2: POST-FURMAN GEORGIA RACE OF VICTIM DISPARITIES,
CONTROLLING FOR THE THREE BARNETT DIMENSIONS¹

	0,1,0 (.00)	0,0,2 (.04)	2,0,1 (.09)		
	W.V. .0 (0/1) B.V. .— ^a	W.V. .07 (1/14) B.V. .0 (0/9) 7 pts.	W.V. .14 (1/7) B.V. .0 (0/4) 14 pts.		
0,0,0 (.00)	1,0,0 (.02)	1,1,0 (.00)	1,1,1 (.25)	1,1,2 (.81)	2,1,2 (.88)
W.V. .0 (0/3) B.V. .0 (0/5) 0 pts.	W.V. .0 (0/13) B.V. .03 (1/30) -3 pts.	W.V. .0 (0/6) B.V. .0 (0/4) 0 pts.	W.V. .34 (14/41) B.V. .05 (1/19) 29 pts.	W.V. .53 (53/64) B.V. .67 (6/9) 16 pts.	W.V. .86 (6/7) B.V. 1.0 (1/1) -14 pts.
	0,0,1 (.00)	0,1,1 (.02)	0,1,2 (.29)	2,0,2 (.56)	
	W.V. .0 (0/30) B.V. .0 (0/25) 0 pts.	W.V. .03 (1/32) B.V. .0 (0/16) 3 pts.	W.V. .38 (6/16) B.V. .0 (0/5) 38 pts.	W.V. .55 (6/11) B.V. .57 (4/7) -2 pts.	
		1,0,1 (.01)	1,0,2 (.26)		
		W.V. .01 (1/89) B.V. .0 (0/95) 1 pt.	W.V. .31 (8/26) B.V. .19 (3/16) 12 pts.		
		2,0,0 (.0)			
		W.V. .0 (0/1) B.V. .— ^a			

¹ Distribution of cases is the same as is figure 1. Each cell indicates the coding on the three Barnett dimensions, the overall death sentence rate (in parenthesis), the death sentencing rates for white victim (W.V.) and black victim (B.V.) cases respectively and the arithmetic difference between the two rates in percentage points.

^a .— indicates no cases.

Arnold Barnett(6), reanalyzed data collected by Baldus, Pulaski and Woodworth for the procedural reform study. Barnett's approach was to read narrative summaries of most of the cases. On the basis of this reading he developed a classification scheme base on three variables: the deliberateness of the killing, the relationship between victim and perpetrator and the vileness of the killing. Figure 2 (Baldus, Woodworth and Pulaski(4), Figure 2) displays 606 cases broken down by Barnett's classification scheme. The triplet at the top of each cell indicates deliberateness (0=not clear that defendant was a deliberate killer, 1=clear that defendant was a deliberate killer, 2=clear that defendant was a deliberate killer, 1=neither), relationship with the victim (0=close relationship, 1=not a close relationship) and vileness of the killing (0=elements of self defense, 2=vile killing, 1=neither). Death sentencing rates for white victim (WV) and black victim (BV) rates are shown. Again, when black victim and white victim cases are compared within matching categories, the death sentencing rate is generally higher for white victim cases, particularly in the moderately aggravated cases where the death sentencing rate is 20 to 30 percent.

Although this is a cursory introduction to studies of race of victim effects, it does show that race of victim disparities persist under a variety of analyses taking numerous background variables into account.

Notes

1. Rosenbaum, Paul R. and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, 1, 41-55 (1983).
2. Finkelstein, Michael O., "The Judicial, Reception of Multiple Regression in Race and Sex Discrimination Cases," *80 Columbia L. Rev.* (1980).
3. Gross, Samuel R. and Robert Mauro, "Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing & Homicide Victimization," *37 Stanford L. Rev.* (1984).
4. Baldus, David C., George Woodworth and Charles A. Pulaski, Jr. "Monitoring and Evaluating Contemporary Death Sentencing Systems: Lessons from Georgia," *18 U.C. Davis L. Rev.* (1985).
5. Baldus, David C., Charles A. Pulaski, Jr., and George Woodworth, "Arbitrariness and Discrimination in the Administration of the Death Penalty: A Challenge to State Supreme Courts," *15 Stetson L. Rev. No. 2* (1986).
6. Barnett, Arnold, "Some Distribution Patterns for the Georgia Death Sentence," *18 U.C. Davis L. Rev.* (1985).

Racial discrimination and arbitrariness in capital punishment: A review of the evidence

Raymond Paternoster
University of Maryland

I think that one somewhat certain observation is that historically racism has figured prominently in the imposition of the death penalty in America. Before the civil war, black codes in the south and border states provided for capital punishment for selected crimes only if committed by black offenders. In 1816 a Georgia statute explicitly provided for the death penalty for the rape of a white woman but only if committed by a black. We also see a pervasive pattern of post-civil war racial discrimination in the imposition of the death penalty. As noted by Bowers in his recent text Legal Homicide since 1864 blacks have been executed more often, for lesser crimes, with fewer avenues of appeal than have whites. In his introduction Colin Loftin mentioned that most capital punishment studies dealing with the issue of racial disparity has been concerned with discrimination by the race of the victim. As a point of clarification it should be said that while most of the recent capital punishment literature has dealt with the race of the victim the earlier literature was concerned with both victim and offender based discrimination. Most of the current studies have also shown a rekindled interest in the effect of race of offender-victim combinations on the imposition of a death sentence. In some of these more recent studies that George Woodworth will discuss we will see the appearance, in particular kinds of cases, of offender-based racial discrimination.

We can document the existence of post-civil war racial discrimination in terms of the race of the offender by looking at the proportion of total state sponsored executions that involved black defendants. Since 1930 with the recording of executions there have been somewhat over 3800 executions of which 53% were of black defendants. This number excludes thousands of illegal executions by lynch mobs conducted against blacks primarily in the South, generally (if not exclusively) for crimes committed against whites. The disproportionate appearance of black offenders in execution data might not reflect racial discrimination, it might simply reflect the fact that black offenders and crimes committed against white victims are more aggravated crimes than those by white offenders or those against black victims, and are therefore more likely to result in a sentence and imposition of death. The central issue for those conducting racial disparity investigations in capital punishment has been the documentation of that disparity and the attempt to explain any such observed differences by legally relevant characteristics of the offense or offender.

Before discussing this literature I would like to draw a distinction between capriciousness and discrimination in the imposition of capital punishment because it will have important implications both in the early and later literature. I would like to define capriciousness as the inability to distinguish with rational criteria between two groups of homicides, a small pool of homicides that result in the imposition of the death penalty and a larger group that does not. We observe the operation of capriciousness or arbitrariness when the former group of cases cannot be rationally distinguished from the latter. Discrimination refers to sentencing or decision outcomes influenced by extralegal status attribute factors such as the race, sex, and social class of the victim or offender. Furthermore, there is an important substantive and legal difference between capriciousness and discrimination. A concern with racial discrimination in capital sentencing appears in the early, pre-Furman literature and first became a constitutional issue via the equal protection clause of the 14th Amendment in McGautha v. California. Capriciousness in capital sentencing became an issue in post-Gregg research where two justices (Stewart and White) overturned states' capital punishment statutes not so much on their being discriminatory in application but rather on the basis of their capriciousness, with Justice Stewart claiming that then existing systems of capital punishment smacked of a lottery system. The issue of capriciousness and discrimination will arise in post-Furman statutes and discrimination exclusively in pre-Furman studies. I would also like to show how the issues of discrimination and capriciousness will eventually converge in that evidence of racial discrimination in capital sentencing is strongest in those categories of homicides that allow for the most discretion and opportunity for arbitrariness. In other words, the cases that have no clear guide will be the ones where racial discrimination will be felt the strongest. To briefly illustrate, we can theoretically construct two kinds of homicides. The first type of case includes those homicides that make up approximately 10% of the overall pool of homicide cases yet 60% of those homicides that result in a death sentence. These are the most egregious, aggravated types of homicides. There is a clear guide for conduct in these kinds of cases, both for prosecutors in deciding to seek the death penalty and for juries in imposing it. There is little discretion for the operation of discrimination in these kinds of cases. The second type of homicide case, however, makes up 90% of the overall pool of homicide cases and 40% of those homicides re-

sulting in a death sentence. In these cases it is not so clear cut that a sentence of death should either be sought or imposed, apparently, they do not cross the requisite threshold of aggravation. There is greater discretion to seek and impose a sentence of death in these cases (capriciousness) and greater opportunity for the appearance of discrimination. What will be seen in our review of empirical studies, then, is the convergence of arbitrariness and discrimination.

The early research on racial discrimination and capital punishment began in the 1940's, with most of the research being conducted with southern states. In 1941 Guy Johnson published research involving the states of Georgia, North Carolina, and Virginia for the period 1939-1940. He examined three stages of homicide prosecution and imposition of the death penalty; (1) indictment, (2) conviction, and (3) sentencing. He found a pattern fairly consistent with other studies in that the probability of a death sentence was highest for blacks who killed whites and lowest for blacks who killed other blacks. Perhaps the most complete and illustrative study was published in 1949 by Harold Garfinkle. A combination of two of Garfinkle's tables shows the movement of capital cases through the North Carolina death sentencing system during the years 1930-1940. Garfinkle focused on three particular decision points; the grand jury's decision to indict for first degree murder, the prosecutor's decision to go to trial on a first degree murder charge, and the jury or trial judge's decision to convict of first degree murder. In North Carolina at that time a conviction of first degree murder carried a mandatory death sentence. The last column of Table 1 summarizes the movement from the beginning to the end of the North Carolina capital punishment process. That column shows the overall probability of a death sentence given a criminal homicide indictment. Looking at the second panel of Table 1, the effect of the race of the offender on the likelihood of a death sentence, there seems to be no effect for the offender's race. The ratio is approximately 1.2 to 1 with white defendants more likely to be sentenced to death. Why the data may not show evidence of racial discrimination against black defendants is that killings of black victims may be very unlikely to result in a death sentence and black offenders are also more likely to have killed a black victim than a white one. In this event race of offender effects would be somewhat obscured without also considering the race of the victim and differential rates of capital punishment for the killing of a white and black. When you control for the race of the victim, offender effects begin to emerge. The top panel of Table 1 indicates that a homicide involving a black offender and white victim has a probability of a death sentence of .29, while a white who kills another white has a probability of .07, a ratio of probabilities of approximately 4 to 1. Consistent with recent literature we also see a strong race of victim effect. The bottom panel of Table 1 (1C) shows that the likelihood of a death sentence in

black victim homicides is .02 while for those involving white victims it is .12, a ratio of 6 to 1. Even in these cases, however, race of victim effects are obscured somewhat by the race of the defendant. One sees this in examining the difference between blacks who killed whites and blacks who killed other blacks. A black killing a white has a probability of a death sentence of .29 in North Carolina during this period, a black killing a black, however, has only a probability of .03, a ratio of almost 10 to 1 when compared with black-white killings. The Garfinkle study is particularly important in showing the existence of racial discrimination, its appearance at different points in the capital sentencing process, and the level of that discrimination in pre-Furman capital punishment statutes.

A problem with the Garfinkle study (and other early studies), however, is that although his category of homicide was somewhat homogeneous (because he examined only first degree murder indictments), these homicides include quite different kinds of acts involving different degrees of aggravation, brutality, and offender culpability. This presents an evidential problem in that the previously discussed evidence of racial disparity in capital sentencing may only reflect the fact that offenses committed against white victims, particularly by black defendants, are more aggravated than those against blacks. The Garfinkle study, and most of this early literature, has very few statistical controls for these legally relevant factors.

Other points in the capital punishment process have been investigated in this early literature. Elmer Johnson in 1957 examined the probability of a commutation of death sentences and found that white defendants were more likely to have their death sentences commuted than black defendants. Like other researchers before him, however, Johnson did not employ statistical controls for other relevant legal characteristics of the homicide he investigated. In 1962, Wolfgang, Kelly, and Nolde examined the execution of black offenders in Pennsylvania, and found that black offenders were less likely to have their death sentences commuted (and therefore more likely to be executed). They also discovered, however, that black defendants were also more likely to have committed felony related homicides, that is, homicides that also involved a contemporaneous felony such as burglary, armed robbery or rape, thereby making it a more aggravated homicide. The effect of the race of the offender still persisted in Wolfgang et al.'s data, however, when they separately considered felony and non-felony homicides.

In the mid-1960's there were two important sets of studies published in the capital punishment literature, one of these was by Wolf (1964) and the other was a series of studies by Wolfgang and his colleagues concerning the punishment of rape in southern states. Wolf examined the sentencing of 159 capital offenders in New Jersey over the period 1937-1961, and consistent with

TABLE 1

Indictments, Charges, Convictions, and Death Sentences in Ten Counties of North Carolina, for Criminal Homicides, by Race of Offender and Victim, from 1930 through 1940

	Numbers at Each State				Conditional Probability of Moving between Successive Stages			
	(1) All Homicide Indict- ments	(2) First Degree Murder Indict- ments	(3) First Degree Murder Charges at Trial	(4) Death Sentences for First Degree Convic- tions	(5) First Degree Indict- ments Given Indict- ments	(6) First Degree Charge Given First Degree Indict- ment	(7) Death Sentence Given First Degree Charge	(8) Overall Prob- ability of a Death Sentence Given Indict- ment
A. Offender/victim racial combinations								
Black kills white	51	48	35	15	.94	.73	.43	.29
White kills white	165	138	73	11	.84	.53	.15	.07
Black kills black	581	531	307	15	.91	.58	.05	.03
White kills black	24	17	8	0	.71	.47	.00	.00
B. Race of offender								
White	189	155	81	11	.82	.52	.14	.06
Black	632	579	342	30	.92	.59	.09	.05
C. Race of victim								
White	216	186	108	26	.96	.58	.24	.12
Black	605	548	315	15	.91	.57	.05	.02

Data Source: Table 7-1; William J. Bowers, Legal Homicide, (p. 208).

most other capital punishment studies he found that blacks were twice as likely as white defendants to be sentenced to death. The importance of the Wolf study is that with the publication of his research we see the introduction of legal controls in estimating the effect of racial influences in the imposition of the death penalty. Wolf controlled for the type of murder, the murder weapon involved in the case, and the offender's age. He found that even with these factors simultaneously considered they did not diminish the observed racial disparity in capital sentencing by the race of the offender. The Wolfgang et al. rape studies (eventually submitted as part of the legal brief in Maxwell v. Bishop examined rape cases in eleven southern and border states for the years 1945-1965. Wolfgang et al. found that blacks who raped whites were 13 times more likely to be sentenced to death than all other racial combinations. In a subsequent reanalysis of the data for Georgia, Wolfgang and Riedel simultaneously controlled for thirteen legally relevant factors, the aggravation of

the homicide and the culpability of the offender. Even with these controls, blacks accused of raping white women were substantially more likely to be sentenced to death than other racial combinations of victim and offender.

This pattern of racial discrimination by race of victim, offender, and combinations of victim's and offender's race continued into post-Furman research. This era of capital punishment research was expanded to include a consideration of both racial discrimination and capriciousness. One of the most important studies in the post-Furman period was the study of Bowers and Pierce of the capital sentencing schemes of Florida, Georgia, Texas, and Ohio. Table 2 reports some of the data reported in their article concerning Florida. These data trace the same process (the movement through the death penalty system) as Garfinkle did in the 1940's, and the similarities in the findings are striking. Bowers and Pierce's research is a significant improvement over

TABLE 2

Charges, Indictments, Convictions, and Death Sentences in Florida for Felony and Nonfelony Homicide, by Race of Offender and Victim (from effective date of post-Furman status through 1977)

Offender/Victim Racial Combinations	Numbers at Each State				Conditional Probability of Moving between Successive Stages			
	(1) All Homicide Charges at Ar- raignment	(2) First Degree Murder Indict- ments	(3) First Degree Murder Convic- tions	(4) Death Sentences	(5) First Degree Indict- ment Given Charge	(6) First Degree Convic- tion Given First Degree Indict- ment	(7) Death Sentence Given First Degree Convic- tion	(8) Overall Prob- ability of Death Sentence Given Charge
Felony								
Black kills white	49	162	74	38	1.000	.456	.514	.234
White kills white	100	208	101	42	.970	.486	.416	.196
Black kills black	35	66	22	7	.800	.333	.318	.085
White kills black	1	10	2	0	-	.200	.000	-
Nonfelony								
Black kills white	18	31	9	1	.722	.290	.111	.023
White kills white	205	249	68	7	.522	.273	.103	.015
Black kills black	279	222	34	4	.312	.153	.118	.006
White kills black	20	10	1	0	.450	.100	.000	.000

Data Source: Table 708, William J. Bowers, Legal Homicide, (p. 293).

Garfinkle's earlier study in that they separated felony from non-felony homicides. The felony homicides were those homicides committed with a contemporaneous felony. Non-felony homicides generally include family slayings, bar-room brawls resulting in a slaying, homicides involving lover's triangles and the like. We can see the same pattern of racial discrimination in Bowers and Pierce's data as in Garfinkle's. The figures in the farthest right column indicate that the ratio of the probability of a death sentence between black-white and black-black killings is 2.75 to 1 for felony homicides and 3.80 to 1 for non-felony homicides. If we collapse offender's race into white victim and black victim homicides we can observe the pattern that was found in the earlier, pre-Furman studies. The probability of an indictment for first degree murder is higher for the slaying of whites than for the slaying of blacks. The probability of conviction is also higher for the slaying of whites than for the slaying of a black. Finally, the overall probability of a death sentence is over twice as likely for those who kill whites than blacks. When we control for at least one other legal consideration (type of homicide involved)

we see a continuation of the racially discriminatory pattern in the imposition of the death penalty. It is clear that the racial effects observed in the more recent Bowers and Pierce study are somewhat diminished from those found in the Garfinkle study simply because Bowers and Pierce separately considered felony and non-felony homicides, whereas Garfinkle's study examined all first degree murder indictments.

The final study that I would like to discuss is the research done by Mike Radelet in 24 Florida counties for the period 1976-1977. Radelet went one step beyond Bowers and Pierce and controlled not only for the type of homicide (looking at only first degree murder indictments, which under Florida's new death penalty statute is the only homicide statutorily eligible for the death penalty) but also for the relationship between the victim and offender, presuming that homicides committed between strangers were more aggravated than those involving acquaintances. Table 3 reports some of Radelet's data, where again there is a higher probability of a death sentence for white victim as opposed to black victim homicides. In non-primary homicides (those involving strangers) the ratio is

TABLE 3
Relationship and Racial Characteristics
of Victims and Defendants for all Homicide
Indictments

	Number of Cases	First Degree Indictments	Probability of First Degree Indictment	Sentenced to Death	Probability of Death Penalty (All Cases)	Probability of Death Penalty (First Degree Indictments)
Nonprimary						
White victim						
Black defendant	63	58	.921	11	.175	.190
White defendant	151	124	.821	19	.126	.153
Black victim						
Black defendant	103	56	.544	6	.058	.107
White defendant	9	4	.444	0	.000	.000
Primary						
White victim						
Black defendant	3	1	.333	0	.000	.000
White defendant	134	73	.545	3	.022	.041
Black victim						
Black defendant	166	51	.307	0	.000	.000
White defendant	8	4	.500	0	.000	.000
N	637	371	.582	39	.061	.105
Nonprimary						
Black defendant	166	114	.687	17	.102	.149
White defendant	160	128	.800	19	.119	.148
White victim	214	182	.850	30	.140	.165
Black victim	112	60	.536	0	.054	.100

Data Source: Table 1, Michael L. Radelet, "Racial Characteristics and the Imposition of the Death Penalty", American Sociological Review, 1981, 46:918-927.

approximately 3 to 1. The consistent evidence reported in the literature of the existence of racial discrimination in capital sentencing does not seem to disappear when you consider at least two kinds of aggravating factors (the type of homicide and the relationship between the victim and offender). Another important factor in Radelet's study was the employment of logistic regression analysis which allowed for simultaneous statistical controls for several non-racial factors. He discovered that killers of whites were more likely to be indicted for first degree murder and were more likely to be given a death sentence overall than were killers of blacks. He also found, however, that once indicted for first degree murder there was no large race of victim effect. Radelet's analysis traces racial discrimination to prosecutorial and grand jury behavior in indicting killers of whites more often than killers of blacks. This is going to be an important consideration in capital punishment research because it locates the point in the system where racial discrimination appears.

In sum, it does appear that there is considerable evidence of victim-based racial discrimination at several points in the processing of potential capital punishment cases. There is inconsistent evidence of offender-based racial discrimination, some studies finding it at particular points but not others, and we see in recent studies precisely where in the system, and for what kinds of homicides discrimination is likely to appear. There are several limitations of the studies discussed thus far, however, which preclude definitive conclusions regarding the role of race in the administration of capital punishment.

The major methodological problem with these studies is that they controlled for only a few of the most important exogenous variables. The observed differences in sentencing and indictment rates for black offenders and killers of whites may be due to the different kinds of homicides committed by black defendants and against white victims; they may be more aggravated, brutal, and may involve more culpable

or violent offenders. These excluded factors can provide a non-racial explanation for the previously reported racial differences. A second problem with these early studies concerns sample selection bias, which will be discussed by Jim Heckman. (Although it should be kept in mind that Radelet found considerable racial discrimination in the screening of cases at earlier stages in the capital sentencing process such as prosecutorial charging and grand jury indictment behavior which is less affected by sample selection bias.) The final limitation of these studies is that they restrict our attention to particular issues to the neglect of others. Most problematic is that they provide us with no theory of the processing of capital cases. They provide no understanding of why particular homicides are likely foci of discrimination, or why victim-based discrimination is more prevalent than offender-based discrimination. I think it is time now to discuss those issues.

The studies presented at this meeting are particularly interesting because they offer so many controls for legally relevant considerations in estimating the effect of race on the imposition of capital punishment (something previous research has not done), and most importantly, because they examine interesting interactions and complexities in their respective datasets.

Barnett's study relied upon an intuitive approach to data analysis in capital punishment research and he is highly critical of the brazen use of multivariate statistics with data in this area. Barnett is at least partially correct in his assessment for several reasons. First of all, these statistical procedures rely on rather strong assumptions about the distribution of the variables and their underlying error structure; assumptions which may not be met by the data. In addition, there is a need to control for many exogenous factors in these causal models which is difficult to do with precision in many kinds of discrete variable multivariate software. Most importantly, this use of multivariate statistics has a tendency to obscure some particularly interesting and kinds of interactions among the variables that are found in the data. In early and more recent capital punishment research we observed the interaction between race of victim and race of offender effects. And there are even more interesting and substantively important kinds of interactions and complexities in recent studies that could be obscured by the uncritical application of multivariate statistical techniques. I do not think you need to be a statistical Luddite to be skeptical of the indiscriminate employment of multivariate data analysis in capital punishment research. Another reason it is important at times to employ simple crosstabulation analyses is in terms of how these data are often used for policy purposes. Many of us who do capital punishment research are called upon at times to

present these findings in court and as one who has attempted to explain what an odds multiplier is to a South Carolina jury I found that particularly difficult to do.

A very simple question need be asked about the studies reported on here; what do these data tell us about the appearance of capriciousness, arbitrariness, and discrimination in the current system of capital punishment? In his study, Barnett classified homicide cases according to three criteria: (1) the deliberateness of the killing, (2) the status of the victim and (3) the brutality or vileness of a particular killing. Barnett's scoring system ranged from a minimum score of zero, the least aggravated kind of homicide, to a maximum score of five. In this scoring system the classification of cases is particularly interesting because it shows rather wide-spread evidence of capriciousness in Georgia's capital sentencing scheme. For instance, in those cases involving scores of two or less there were only three death sentences out of 373 (a probability of .008). In the category scored three by Barnett, there were 33 out of 134 defendants that received a death sentence (a probability of only .25). In the highest category (scores 4 or 5) there were 76 death sentences out of 99 cases (a probability of .77). In those cases at the highest level of aggravation there was a 77% death sentencing rate which may be looked at in terms of the glass being 3/4 full or 1/4 empty. Although the rate of capital sentences is high in these kinds of homicides the data indicate that about one in four of the most aggravated of homicides does not result in a death sentence while three in four factually similar cases it does. More disturbing than this, though, is that in Barnett's category three homicides, which make up 22% of all capital crimes and 29% of all death sentences, the imposition of the death penalty is a particularly capricious event, occurring only in about 25% of those cases.

A similar kind of evidence of capriciousness in death sentencing is found in Gross and Mauro's study and that by Baldus, Woodworth and Pulaski. Gross and Mauro examined (principally) the capital sentencing schemes in Georgia, Florida and Illinois. They constructed a homicide aggravation scale that ranged from zero to three. The aggravating factors of a homicide their scale considered were; (1) the existence of a contemporaneous felony, (2) the relationship between the victim and offender, and (3) the number of victims. These data are presented in Table 4 where it can be seen that there is a reasonable degree of consistency in the pattern of capital sentencing. As the level of aggravation of the homicide increases the probability of a death sentence increases. These data clearly suggest, then, that there is some rationality in the system. However, it also can be seen that even at the highest levels of aggravation in these three states the actual imposition of the death penalty is less than a

TABLE 4

Percentage of Death
Sentences by Level of
Aggravation

	Number of Major Aggravating Factors			
	0	1	2	3
GA	0.4% (6/1635)	7.7% (26/339)	31.6% (43/136)	57.1% (4/1)
FL	0.6% (14/2295)	4.7% (41/874)	21.9% (62/283)	44.0% (11/25)
IL	0.1% (2/1924)	1.0% (7/711)	7.4% (29/392)	22.6% (7/31)

Data Source: Table 21, Samuel R. Gross and Robert Mauro, "Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing and Homicide Victimization", 37 Stanford Law Review (1984).

moderate certainty. Looking at categories one and two, where most of the homicides are classified, it is for the most part quite unlikely that a death sentence is going to be imposed; evidence of the capriciousness of capital sentencing even under a reformed post-Gregg statute.

Table 5 presents a table from the Baldus, Woodworth and Pulaski study in which Georgia homicides are classified on the number of statutory aggravating factors present, and again there is some degree of consistency and rationality in the system in that the greater the number of statutory aggravating circumstances appearing in the case, the greater the likelihood of a death sentence. It is also quite clear, however, that in the middle range of aggravation where most of the homicides are classified (categories one, two, and three) the probability of a death sentence is less than 50 percent. Further evidence of capriciousness appears in their regression based scales where Baldus, Woodworth and Pulaski estimate the likelihood of a death sentence and then collapse the cases into levels of aggravation from 1 to a high of 8 (see Table 6). A recognized pattern emerges in this analysis where only at the highest levels of aggravation is there a fairly substantial likelihood of a death sentence being imposed. Again, for the three categories where most of the homicides are found the probability of a death sentence is less than one-half.

In sum, it is clear that even in these post-Gregg statutes there is still considerable evidence of capriciousness in sentencing, and that for a great many homicides (most of them in the overall homicide pool) the likelihood of

TABLE 5

Death Sentencing Rates Controlling
for the Number of Georgia Statutory
Aggravating Factors Present

A	B
Number of Statutory Aggravating Factors Present	Death Sentencing Rate
0	.0 (0/132)
1	.03 (5/150)
2	.12 (16/136)
3	.37 (37/99)
4	.53 (33/62)
5	.62 (8/13)
6	.50 (1/2)

Data Source: Table 2, David C. Baldus, Charles Pulaski, and George Woodworth: "Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience," Journal of Criminal Law and Criminology 74:661-753.

a death sentence is very small. The importance of this point is that capriciousness and discrimination in capital sentencing converge in these data. Racial discrimination will appear strongest at middle levels of homicide aggravation. Table 7A reports data from Barnett's study where the relationship between the likelihood of a death sentence and race of victim is shown in the two categories of cases where the death penalty was at least a statutory possibility (his category three and four). In category three the probability of a death sentence is very small and contains about 55 percent of all capital crimes in the pool (Georgia data). The likelihood of a death sentence only increases to a high level in more aggravated capital crimes, but even here in the most aggravated homicides there are subtypes where the death penalty is only a 50-50 probability. It is in the former cases, the category 3 homicides, where racial discrimination

TABLE 6

Race of Victim Disparities in Death Sentencing Rates Among Defendants Indicted for Murder
Controlling for the Predicted Likelihood of a Death Sentence and Race of the Victim

<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>	<u>G</u>	<u>H</u>	<u>I</u>	<u>J</u>
Predicted Chance of a Death Sentence 1 (least) to 8 (highest)	Average Actual Sentenc- ing Rate for the Cases at Each Level	Death Sentenc- ing Rates for Black Defense Involving White Victim Cases	Black Victim Cases	Arithmetic Difference in Race of the Victim Rates (Col.C/ Col.D)	Ratio of Race of the Vic- tim Rates (Col.C/ Col.D)	Death Sentenc- ing Rates for White Defend- ant Involv- ing White Victim Cases	Black Victim Cases	Arithmetic Difference in Race of the Victim Rates (Col.G- Col.H)	Ratio of Race of the Vic- tim Rates (Col.G/ Col.H)
1	.0 (0/33)	.0 (0/9)	.0 (0/19)	.0	0	.0 (0/5)	-	-	-
2	.0 (0/55)	.0 (0/8)	.0 (0/27)	.0	0	.0 (0/19)	.0 (0/1)	.0	0
3	.08 (6/76)	.30 (3/10)	.11 (2/18)	.19	2.73	.03 (1/39)	.0 (0/9)	.03	0
4	.07 (4/57)	.23 (3/13)	.0 (0/15)	.23	-	.04 (1/29)	-	-	-
5	.27 (15/58)	.35 (9/26)	.17 (2/12)	.18	2.06	.20 (4/20)	-	-	-
6	.17 (11/64)	.38 (3/8)	.05 (1/20)	.33	7.60	.16 (5/32)	.50 (2/4)	-.34	.32
7	.41 (29/71)	.64 (9/14)	.39 (5/13)	.25	1.64	.39 (15/39)	.0 (0/5)	.39	0
8	.88 (51/58)	.91 (20/22)	.75 (6/8)	.16	1.21	.89 (25/28)	-	-	-

Data Source: Table 42, David C. Baldus, George Woodworth and Charles Pulaski, Discrimination and Arbitrariness in Georgia's Capital Charging and Sentencing System. A Preliminary Report, unpublished monograph.

is strongest. The probability of a death sentence is almost twice as high for white victim as black victim homicides for category three killings, while the difference is negligible for more aggravated, category 4 homicides. The same is true in Gross and Mauro's aggravation scale (Table 7B). At the highest level of aggravation there appears to be little evidence, or only very limited evidence, of racial disparity in capital sentencing. In the middle range of aggravation, however, where most of the homicides occur, there is indeed very substantial evidence of victim-based racial discrimination.

It appears that racial discrimination perhaps does not pervade the entire system of capital sentencing—just a very substantial part of it. The data appear to indicate that there exists a threshold in homicide cases. There is some consistency and rationality in the system in that as a threshold of aggravation is reach-

TABLE 7A

Probability of A Death Sentence by Race of
Victim for Barnett's Category 3 and 4
Homicides

	Black Victim	White victim
Category 3	.11	.32
Category 4		
(1,1,2)	.75	.84
(2,0,2)	.57	.55

Data Source: Arnold Barnett, "Some Distribution Patterns for the Georgia Death Sentence," 18 U.C. Davis Law Review (1985).

TABLE 7B

Probability of a Death Sentence by Level of Aggravation from Gross and Mauro Study

	Level of Aggravation				0 cases included	0 cases excluded
	0	1	2	3		
White Victim	.008 (4/499)	.101 (18/179)	.456 (41/90)	.800 (4/5)	.087 (67/773)	.230 (63/274)
Black Victim	.002 (2/1136)	.050 (8/160)	.043 (2/46)	- (0/2)	.009 (12/1344)	.048 (10/208)
Ratio of Probabilities	4.00	2.02	10.60	-	9.67	4.79

ed and crossed the probability of a death sentence increases for white and black victim homicides. More disturbing though is the fact that in some categories of cases different sentencers operate with different thresholds such that a given homicide may cross the requisite aggravation threshold for one sentencer, and a death sentence is imposed, but does not cross the threshold for other sentencers, and the death penalty is not imposed. This is the process that appears to be creating arbitrariness in post-Gregg capital punishment statutes.

Third, and perhaps most disturbing, is that it appears that a different threshold of aggravation exists for the killing of whites and blacks. The data tend to indicate that the probability of a death sentence becomes equivalent for white and black victims only at the highest levels of aggravation. For white victim cases there is a greater likelihood of a death sentence than for black victim cases at lower levels of aggravation, indicating that southern juries may be operating with a race-specific definition of homicide severity or aggravation. They appear more willing to tolerate a homicide involving a black than a white victim. At the end of this discussion I will try to suggest an explanation for these findings that does not rely on the racism of individual actors within the criminal justice system. Before that, however, it will be informative to return to Barnett's work for just a moment.

As capital punishment researchers we are above all else trying very hard to explain capital sentencing behavior, generally the behavior of juries, as Barnett attempts to do with his classification scheme. His classification scheme does indeed indicate that at high levels of aggravation (his category 4) Georgia juries behave consistently. I would like to suggest that just because Georgia juries behave consistently does not necessarily mean that they be-

have rationally or fairly. To illustrate what I am suggesting, I would like to show you the importance of Barnett's third criteria in his classification scheme (whether or not the offense is vile or heinous). That third element is particularly important in escalating a case to a death sentence. We can construct two homicides according to Barnett's classification scheme that differ only in that one of them meets the criteria of a "vile and heinous" homicide. This difference is enough to increase the likelihood of a death sentence fairly substantially (a factor of about 3 (see Table 8). One of the pieces of information used in classifying a case as vile or heinous in Barnett's scheme is the description of the crime scene found in the homicide narrative for each case. A graphic crime scene that is bloody and gruesome is included in Barnett's

TABLE 8

Increase in the Probability of A Death Sentence Due to the Inclusion of Criteria 3 in Barnett's Classification Scheme (the brutality of a Homicide)

Homicide Category	Probability of a Death Sentence
1, 1, 1	.27
1, 1, 2	.35
2, 0, 1	.17
2, 0, 2	.53

scheme as fitting his category three criteria, and is predictive of Georgia jury sentencing behavior. Georgia juries do seem to employ this information in determining who should live and who should die such that we can use it to explain their behavior. This does not necessarily mean, however, that they should be using that criteria; in fact, there is rather substantial evidence that they should not be.

The U.S. Supreme Court in Godfrey v. Georgia has indicated that what is important is not the understanding of jury behavior, but that jury behavior should be rational. The Court has consistently noted that there should be a rational basis for distinguishing those cases that result in a death sentence from the larger number of cases that do not. The appearance of a crime scene, the actual physical appearance of the crime scene is not such a rational basis. In Godfrey v. Georgia just such a situation arose. I scored Godfrey's case according to Barnett's scheme and scored it a four, a highly aggravated homicide. This particular case involved a defendant who after several days of argument with his estranged spouse went to her mother's trailer and killed both his wife and mother-in-law, and assaulted his daughter as she fled the scene. The killing was achieved with a shotgun, a weapon the Court noted not known for its surgical precision, and the narrative of the case describes the crime scene as horrible, in colorful terms, and it certainly would fit Barnett's classification scheme as a "one, one, two" homicide wherein 85% of the defendants are given a death sentence.

It would seem, on the face of it, that the Godfrey case is not a particularly good example of capriciousness in sentencing except that the U.S. Supreme Court in Godfrey v. Georgia overturned Godfrey's death sentence on the basis that the jury acted irrationally. The majority opinion noted that there was very little to distinguish Godfrey's killing from the hundreds of other spousal killings in Georgia, except for the brutal appearance of the crime scene, which, the Court said the jury should not have used in determining Godfrey's sentence. The point is that we as capital punishment researchers should be careful as to the criteria we employ and information we use in our multivariate models explaining jury behavior. To determine those factors that statistically explain jury sentencing behavior is not at all to say that juries behave fairly or rationally, in spite of the high R^2 's of our models.

As a final point let me make another comment on the brazen employment of multivariate statistical techniques and how their indiscriminate use may hide particularly interesting kinds of interaction effects in the data. To repeat an earlier point, it was noted that the appearance of racial discrimination is highest in middle-level aggravation cases. When these cases are aggregated into one pool of homicide cases we are combining types of homicides where the racial effect may be diminished at the two ends of the spectrum and very high in the middle range, such that the overall evidence of racial

disparity in sentencing is muted. There is other evidence of interaction effects being potentially hidden with the use of multivariate statistical techniques. For instance, in the Baldus, Woodworth and Pulaski data set they found that where a B2 aggravating circumstance is found (in Georgia, a B2 aggravating circumstance is the commission of another statutorily specified felony in addition to homicide) a monetary motive as an additional aggravating circumstance has no effect on sentencing, but in the absence of a contemporaneous felony a monetary motive for homicide has a strong effect on the probability of a death sentence. Their preliminary multiple regression analyses obscured this but they did find and report this observation in their crosstabulations. I think that their finding is not only methodologically but substantively important in that it may indicate the existence of "charge stacking" on the part of prosecutors. If prosecutors have evidence of an armed robbery they may also likely charge the defendant with an additional aggravating circumstance of homicide for pecuniary gain. In doing so they tend either to enhance their position in plea bargaining, or successfully "upgrade" an offense making it appear more serious in its presentation to a jury.

Baldus, Woodworth and Pulaski also found that when there was a contemporaneous felony the impact of Georgia's B7 statutory aggravating circumstance (whether or not the offense was vile and heinous) is very strong. In the absence of an additional felony, the effect of a B7 aggravating circumstance was substantially reduced, indicating both the practice of charge stacking and that perhaps (in the aftermath of Godfrey) a B7 circumstance in and of itself may not be enough to push a homicide past the threshold of aggravation implicitly employed by capital juries. Finally, I think some of these simple crosstabulations effectively tease out substantively important race of victim and race of offender disparities. We can see that under particular circumstances strong race of victim effects emerge, for example, among those offenses involving contemporaneous felonies, which also happen to be those homicides which are most likely to result in a death sentence. Strong race of victim disparities also arise when there is only a single victim. Strong race of offender effects arise in a Georgia B7 homicide, the vile heinous offense wherein the sentencing standard is somewhat vague. Race of offender effects are also found in rural areas of the state. The data do seem to show clear race of victim and offender effects in capital sentencing that become more pronounced under particular conditions. For these effects simple crosstabulation analysis is particularly important in discerning the underlying relationships.

To close, let me make some suggestions for future capital punishment researchers. First, I think we need to focus on particular kinds of offenses in our analyses. For instance, those involving contemporaneous felonies, particularly those involving armed robbery which, at least in South Carolina the data with which I am most familiar, are a particular focus of racial discrimination. We also need to focus greater energy on particular locations where discrimination may appear, both in terms of geographic location within a particular state and the location of a decision within the capital sentencing system itself. One location of particular importance that Jim Heckman discussed is criminal justice processing in rural areas. Two of the studies discussed today find considerable racial discrimination in rural but not urban areas. All of them find particularly strong evidence of racial discrimination in the prosecutor's charging decision, and whether or not a case was sent to the penalty phase of the bifurcated trial (in Baldus and Woodworth's Georgia data). Finally, I think we need more theoretical development in the area. As I stated before, we are not particularly insightful in explaining why race of victim effects emerge in the first place, and why they dwarf race of offender effects in importance. We need to develop a theory of prosecutorial behavior and one which is not exclusively dependent on the racism of individual actors or the criminal justice system itself. Let me just suggest one for discussion.

We can develop a theory of prosecutorial behavior which can explain race of victim effects not by relying on the racist attitudes of prosecutors but with reference to their behavior in two other roles. One of these

roles is as an advocate in that the prosecutor is compelled to press complaints made by victims' family or a generally expressed sense of community outrage at the offense. The second role occupied by local prosecutors is that of an administrator where they are interested in maximizing their available resources. With respect to the first of the prosecutor's roles it is clear from public opinion polls that whites are more in favor of capital punishment than blacks. This may induce white members of the community to push harder for local prosecutors to seek a death sentence when a white is a victim of a homicide than for black victims. Both empathy for and identification with white victims may lead to more vocal support for (and more likely imposition of) a death sentence for killers of whites. Such identification with the victim is less likely for the killers of blacks, which, when coupled with diminished support for capital punishment generally among blacks, leads prosecutors not to seek a sentence of death, and white dominated juries less likely to impose one. Prosecutors, being concerned about two particular things--one, not wasting resources and two, maximizing the likelihood of a conviction and requested sentence--may be more inclined to seek the death penalty in white victim cases, and may try to enhance that likelihood by making them appear to be more serious than the killing of blacks. Through this process race of victim effects would arise, but as a response to differential pressures and demands on the prosecutor. In any event, this is mere speculation, but it is important that more of such speculation and research continue.

**Comments on three studies of disparity in capital sentencing
by race of victim: Baldus, Woodworth, and Pulaski;
Gross and Mauro; and Paternoster**

James J. Heckman
University of Chicago

These papers represent excellent applications of conventional statistical methods to the analysis of an important social problem. Without doubt, these studies establish the existence of an important race of victim statistical regularity in capital sentencing rates. The thoroughness of these studies and their candor set a high standard for research in legal statistics.

While I have no serious quarrels with the main facts presented, I have some difficulty with the interpretation to be placed on them and their value in any specific case, especially *McCleskey vs. Kemp*. Before presenting my reservations, I will summarize the five features of the data that clearly emerge from these studies.

(1) When a black kills a white, the defendant is much more likely to receive a death sentence than if a black kills a black or a white kills a white. (2) The event "white kills black" is a very rare event. Interracial murder is almost invariably "black kills white". (3) When a black kills a white, there are more aggravating circumstances than in other types of murder (see Table 22 of Gross and Mauro). "Black kills white" murders are rarely domestic violence murders. (4) As Table 1 suggests, the differential capital sentencing of blacks who kill whites is most pronounced in rural areas. The race of victim effect is much weaker in urban areas. (Table 1, from Baldus, Woodworth and Pulaski). (5) None of the studies has adequate data on community response to the murders, the rarity or prevalence of murder of any kind in the community and the relative (to the community) status of the victims.

What inference about discrimination can be drawn from these facts, or for that matter, from any competent statistical study? Very few, if any, without a clear understanding of how the law is supposed to operate if it is nondiscriminatory. Implicit in these studies that do not control for community perception effects of crimes is the view that the jury system should act uniformly across jurisdictions within or across states. Perhaps the jury system should act in such a fashion but nothing in the law requires this. A heinous crime in one location may be an ordinary event in another. Differential responses to identical facts is almost guaranteed by the peer jury system. Since no study has quantified the relative impact of the crimes on the community, none controls for a legitimate variable. (Point 5). The fact that disproportion in sentencing is found most strongly in disparate rural areas where murder rates are low and the crime of murder is a very unusual event (Point 4) reinforces this point. Disparity as measured may not mean discrimination according to the law. Evidence of consistent patterns of disparity across states of the sort presented by Gross and Mauro may merely indicate that the same

sorts of community relative status variables have been left out of all of the studies. (Is it the same type of overt discrimination is operating in Georgia as in Illinois or the same sort of omitted community variables?)

It is unfortunate that most of these studies focus on Southern states. If there were comparable studies on Northern states (or states with populations less likely to discriminate against blacks) that displayed the same type of race of victim effect, the discrimination interpretation of the evidence would be less plausible unless, of course, it is assumed (with no evidence at all) that discrimination is identical in all regions of the U.S.

Putting aside the issue of uniform treatment, all of these studies can be subject to the obvious criticism that some variables relevant to the case and known to the judge and jury are omitted from the statistical analysis. The fact that this is such an obvious objection does not render it invalid. In light of point (3) above, there is considerable reason to doubt that all of the aggravating nuances in these cases have been recorded.

Also missing from all of these studies is a suitable benchmark for measuring fairness. Granting for the moment that all of the relevant community variables are properly measured, what is a fair capital sentencing system? It is one with no predictability? Would that be a capricious or a fair system? Is it one with perfect predictability? An automatic "objective" rule would surely violate the law as recent decisions on North Carolina laws make clear.

In view of point (4), the relevance of these studies to *McCleskey vs. Zant* is less than obvious because the crime that initiated that case occurred in urban Fulton County Georgia.

Some Methodological Points

(1) The Baldus-Woodworth-Pulaski study is for a sample of people who are arrested and convicted. By conditioning on an outcome of the criminal justice system, perverse findings may be produced. Suppose that courts are bending over backward to avoid prosecuting blacks who kill whites. Then only heinous cases will show up in convicted samples. With some unobserved (by the statistician but not by the actors in the legal system) characteristics relevant to the case, the data may still show discrimination against blacks by race of victim solely as a consequence of selecting a sample on the basis of an outcome (arrest and conviction). The Gross and Mauro study is much less vulnerable to this criticism because the primary unit of analysis is a documented homicide. Nevertheless, even this study is not entirely clean if local law enforcement efforts are devoted to doc-

TABLE I

UNADJUSTED URBAN AND RURAL DEATH SENTENCING RATES AND RACIAL
DISPARITIES AMONG DEFENDANTS INDICTED FOR MURDER

I. Death Sentencing Rates Controlling for Defendant/Victim Racial Combination

	Black Defendant/ White Victim	White Defendant/ White Victim	Black Defendant/ Black Victim	White Defendant/ Black Victim
A. Urban (n=1226)	.13 (14/110)	.09 (27/303)	.01 (8/778)	.03 (1/35)
B. Rural (n=1107)	.31 (36/115)	.07 (31/405)	.02 (10/564)	.04 (1/23)

II. Race of Victim Disparities

A. <u>Race of Victim</u>	Urban Circuits <u>Race of Defendant</u>		Rural Circuits <u>Race of Defendant</u>	
	<u>Black</u>	<u>White</u>	<u>Black</u>	<u>White</u>
1. White Victim	.13	.09	.31	.07
2. Black Victim	.01	.03	.02	.04
a. Difference	12 pts.	6 pts.	29 pts.	3 pts.
b. Ratio	13 to 1	3 to 1	16 to 1	1.8 to 1
c. Overall Measure ^{1/}	.10 (.0004)		.25 (.0001)	

III. Race of Defendant Disparities

A. <u>Race of Defendant</u>	Urban Circuits <u>Race of Victim</u>		Rural Circuits <u>Race of Victim</u>	
	<u>Black</u>	<u>White</u>	<u>Black</u>	<u>White</u>
1. Black Defendant	.13	.01	.31	.02
2. White Defendant	.09	.03	.07	.04
a. Difference	4 pts.	-2 pts.	24 pts.	-2 pts.
b. Ratio	1.4 to 1	.33 to 1	4.4 to 1	.50 to 1
c. Overall Measure ^{1/}	.02 (.43)		.18 (.0001)	

^{1/} The overall measures are the regression coefficients for the racial variables estimated with no background controls for non-racial factors.

umenting "heinous" crimes (as perceived by the community), such crimes overrepresent blacks killing whites, and not all of the data relevant to the case is known to the legal statistician.

(2) All of the studies ignore the correlation across observations due to common judges and origins of juries. This uncorrected correlation biases the reported test statistics. There is a modest presumption that it biases reported statistical significance levels upwards (and therefore in favor of finding racial disparity). In addition, none of the studies accounts for variation

in outcomes by jurisdiction despite the fact that a technology for doing so exists (random coefficient models). Evidence of no jurisdictional effects would bolster the conclusions of these studies. In view of fact (4), I doubt that such a conclusion can be drawn.

(3) The very interesting interactions detected by the simple but robust cross classification analysis reported at the end of the Baldus-Woodworth-Pulaski study is very enlightening. It revealed to me the important role of rural location in generating the race of victim finding. Such evidence

casts doubt on the validity of conventional multivariate analyses widely used in legal statistics that ignore such interactions entirely or impose strong restrictions on the nature of admissible interactions.

(4) Following up on remark (3), I am troubled by legal scholars who make frequent appeals to a non-existent statistical authority about how to build a statistical model. The fact of the matter is that there is no objective "best" way to build a statistical model up from a set of data. Despite claims to the contrary in elementary statistics and econometrics books, there is much current controversy over this topic in the professional literature. Conventional pre-test procedures used by many of the authors (*i.e.* include a variable if its associated coefficient has a "big enough" "t" ratio) have no formal justification. The conclusions of these studies would be much more plausible if nonparametric methods were used such as those developed by Breiman, Friedman, Morgan,

Olshen, Sondquist, Stone and others. (See, *e.g.* Breiman et. al, Regression and Classification Trees, Wadsworth, 1983). It will be valuable to see if the race of victim effect holds under more general types of analyses and to find out what configurations of the data give rise to the race of victim effect.

Papers Cited

- Baldus, Woodworth and Pulaski, "Discrimination and Arbitrariness in Georgia's Capital Charging and Sentencing System: A Preliminary Report", unpublished manuscript, University of Iowa, 1983 (7/29/83).
- Gross, S. and R. Mauro, "Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing and Homicide Victimization", 37 *Stanford Law Review* (1984).
- Paternoster, R. "Race of Victim and Location of Crime", Journal of Criminal Law and Criminology, Vol. 74, 1983.

A reply to Professor Heckman

Samuel R. Gross
University of Michigan

Professor Heckman begins his comments by praising the studies he reviews. The papers are "excellent applications of [statistical analysis to] an important social problem;" their thoroughness and candor "set a high standard for research in legal statistics." But praise is no substitute for understanding, and, unfortunately, the discussion that follows reveals some fundamental failures on that score.

Heckman lays the basis for his comments by listing several "features of the data that emerge clearly from these studies." His list is a mixed bag:

(1) & (2) Any person who kills a white victim is more likely to be sentenced to death than those who kill blacks, and there is some (weaker) evidence that blacks who kill whites are at greater risk than whites who kill whites. True; these are the inescapable conclusions of the studies.

(3) Most interracial killings involve white victims. White-kills-black homicides are rare. This is also true, but apparently irrelevant to the discussion. In any event, if this pattern has some significance it is never mentioned.[1]

(4) As a group, homicides with black killers and white victims are more aggravated than homicides with other racial combinations. This too is true, although the comparison to the other cross-racial category--the rare white-kills-black homicides--is not reliable. It is also true, although Heckman does not note it, that white-victim homicides in general are more aggravated than black-victim homicides--at least as the criminal justice system keeps track of such things. Needless to say, the main goal of these studies is to control for these differences in levels of aggravation.

(5) "The differential capital sentencing of blacks who kill whites is most pronounced in rural areas. The race-of-victim effect is much weaker in urban areas." These statements are confusing, and at least partly false. Heckman seems to imply that the "differential sentencing" of blacks who kill whites is the only race-of-victim effect. This is not so, as he has noted. More important, it is not true that the race-of-victim effect is much weaker in urban areas. In our own study we checked for this in one set of the many regression analyses that we conducted, and found that controlling for the urban or rural location of the homicides had approximately no effect on our race-of-victim coefficients (Gross & Mauro, p. 82). Baldus, Woodworth and Pulaski made similar findings in Georgia: after controlling for many other variables in any number of different ways, the race-of-victim disparities that they observed were essentially unaffected by

the location of the homicides. Baldus et al. did find that the smaller and weaker race of defendant effect that they detected was primarily restricted to rural counties. Perhaps this is what Heckman had in mind--it is not clear--but it is quite a different point from the one he states.

(6) The studies do not have "adequate data" on community responses to murders, the prevalence of different types of in the communities at issue, and the community status of the victims. I am puzzled by these statements. I'm not sure I understand Professor Heckman's point, but he seems to be at least partially in error. Each of the studies has detailed information on the patterns of different types of homicides within the geographical units in which these homicides are considered. Thus, for example, Robert Mauro and I have information on the number of felony-circumstance homicides, multiple homicides, etc., in each county in each of the states that we studied, and data on the number of death sentences and their circumstances from each county in each state. Baldus, Woodworth and Pulaski have vastly more detailed information on the various types of homicides that occurred in every county in Georgia. Moreover, this research was conducted and reported against a background of decades of criminological research on homicide patterns in the United States, and there is nothing remarkable about the distribution of homicides or their characteristics in the jurisdictions covered by these studies.

On the basis of this list of "facts," Professor Heckman concludes that "very few [inferences], if any" about discrimination can be drawn from the studies. The links between the initial list and the final statement are not all plain, but the core of the argument seems to be as follows:

The studies are faulty because they fail to "control for community perception effects of crime," and implicitly assume "that the jury system should act jurisdictionally uniformly within or across states."

The second part of this argument is simply false. Robert Mauro and I did not assume any jurisdictional uniformity; we examined each state separately, and the uniform presence of a race-of-victim effect in each state emerged from the data. We also did our best to control for geography within states; David Baldus and his colleagues did much better at that in Georgia. Heckman goes on to state that "nothing in the law requires this" type of jurisdictional uniformity. This assertion is in part debatable and in part false. Several of the states that we examined--conspicuously, Georgia and Florida--have state-wide systems of "proportionality review." One of the purposes of each such system is to enable an

appellate court with "state-wide jurisdiction" to ensure even-handed application of death penalty statutes across the entire state. (See, e.g., Gross & Mauro at pp. 83-85.) Moreover, even in the absence of these explicit provisions, it is possible that other provisions of state and federal law may require at least some types of geographic uniformity in the application of all penal statutes, and especially death penalty laws.

Heckman's argument on "community perception" is equally incorrect. In what is undoubtedly an exercise in hyperbole, he says that "a heinous crime in one location may be an ordinary event in another." If the crime is homicide, that is, of course, entirely false. What is considered a heinous homicide in one location might be considered a somewhat less heinous homicide in another location, but that is all. One of the major lessons of the studies reviewed here, and of a large body of research apart from these studies, is that on the whole people across jurisdictions use similar factors to grade the severity of homicides. Their responses are not identical--far from it--but a reasonable degree of uniformity in the grading of homicides is both well established empirically, and a predictable consequence of our common culture and common humanity.

At root, Professor Heckman seems to have his mind set on two points. First, he is misled by his mistaken view that racial discrimination in capital sentencing is essentially a rural phenomenon. This premise leads him to speculate that "community differences"--such as the differences between urban and rural communities--might somehow explain these racial patterns. But the premise is false, as I've explained.

Second, Professor Heckman seems to believe that "discrimination" means conscious and deliberate bigotry. He does not say that in so many words, but that seems to be the drift of his comments, especially his observation that a consistent finding of similar race-of-victim effects in northern as well as southern states would undercut the claim that this effect reflects "discrimination." But discrimination is not restricted to the worst and most explicit forms of racism, and racism is not restricted to the South. Our study does indeed find racial discrimination in capital sentencing both in the South and outside the South,[2] and, Professor Heckman notwithstanding, this pattern fortifies our confidence in our conclusion. These findings do not require us (as Heckman asserts) to assume "that discrimination is identical in all regions of the country;" they do not show identical patterns of discrimination. They do show--unsurprisingly--that the problem of racial discrimination is present across the United States.

Professor Heckman is a distinguished scholar and an accomplished methodologist. But his unquestioned competence and intelligence are not proof against the type of errors that are hard to avoid when a large body of empirical research is reviewed without a sufficiently detailed study of its contents and its context.[3]

Notes

1. It might be worth noting one of the major explanations for this asymmetry, although it is orthogonal to the focus of these studies. In a segregated society, interracial homicides are almost inevitably homicides between strangers.

Other things being equal, a white who goes outside his circle of acquaintances to commit a homicide is unlikely, on purely statistical grounds, to hit a black victim, or any other minority victim. By contrast, a black who goes outside his circle of immediate acquaintances to commit a homicide will more quickly run into white victims, since whites are the majority. If we assume that in all homicides of strangers the victims are chosen at random, the vast majority of interracial homicides would involve white victims.

2. The Gross and Mauro study examines capital sentencing patterns in Oklahoma, Illinois, and several Southern states.

3. Professor Heckman makes other errors that I have not mentioned, but they are less central. One, however, requires a comment. He states that the relevance of these studies to the McCleskey case is "less than obvious because the crime that initiated the case occurred in urban Fulton County Georgia." Apparently, he is restating his mistaken assertion that race-of-victim effects occur only in rural counties. In this case, however, the error is more specific. Baldus and his colleagues conducted a separate analysis of the Fulton County subset of their data for the McCleskey hearing, and they found the same patterns of racial discrimination that they found in their state-wide analyses.

References

- Baldus, David C., George Woodworth, and Charles A. Pulaski, Jr., "Discrimination and Arbitrariness in Georgia's Charging and Sentencing System," preliminary report (1983).
Baldus, David C., George Woodworth, and Charles A. Pulaski, Jr., "Discrimination and Arbitrariness in Georgia's Charging and Sentencing System," monograph in preparation (1986).
Gross, Samuel R., and Robert Mauro, "Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing and Homicide Victimization," 37 Stanford L. Rev. 27 (1984).

Reply to Professor Heckman

George G. Woodworth, David C. Baldus, and Charles A. Pulaski, Jr.
Arizona State University

Professor Heckman raises four questions which call for a reply. They concern: (a) omitted variables, (b) the strength of race of victim disparities in urban and rural areas, (c) the relevance of our Georgia research to the constitutional claims of Warren McCleskey, and (d) methodological issues deriving from possible correlation across observations due to common judges and juries and the need to account for variations in outcomes by jurisdiction.

Professor Heckman faults our Georgia research and other similar capital punishment studies for failure to control for "community response to the murders" and the "relative impact of the crimes on the community." We assume that this community response variable refers to the possibly different values of various communities that determine how punitively a community will react to a given homicide. His argument appears to suggest the possibility that the white-victim cases may occur primarily in communities which have a generally more punitive reaction to homicide, and that the black-victim cases may be concentrated in the communities which generally have a less punitive response to homicide. Under these circumstances, we could observe a statewide race-of-victim effect even though black- and white-victim cases were treated similarly within the local judicial circuit, which defines the scope of prosecutorial jurisdiction. This is a legitimate concern which falls under the rubric "Simpson's paradox," and we made a specific effort to deal it in the 1983 report of our Georgia research (the Charging and Sentencing Study (CSS)) referred to by Professor Heckman. [1]

The problem in dealing with the issue is how to measure the community response to murder. There are no public opinion data available which would provide a readily quantifiable measure of those differential attitudes. However, we consider a proxy measure quite suitable for handling this problem. Indeed, we consider it preferable because it indicates the "revealed preference" of the prosecutors and jurors in each judicial circuit. Specifically, we created a separate variable for each judicial circuit, each of which, as noted above, falls under the jurisdiction of a single district attorney. We then conducted a multiple regression analysis which estimated a statewide race of victim effect after controlling

simultaneously for those 42 geographic variables and a number of variables for legitimate aggravating and mitigating case characteristics. Six circuits emerged as having unusually high death-sentencing rates and two had particularly low rates. The presence of background controls for these geographic variables did not, however, affect the magnitude or statistical significance of the race-of-victim disparities estimated statewide. The results of this analysis were reported in the 1983 CSS report. [2] Since 1983, we have conducted similar analyses using the data from the parallel Procedural Reform Study (PRS) which controlled for more background factors, as well as for (circuit x race of defendant) and (circuit x race of victim) interaction terms, and the statewide race of victim disparities persisted in all of these analyses just as they did in the CSS analyses described above. [3]

Professor Heckman also faults our failure to control for "the relative (to the community) status of the victims." We know of no way to measure the status of victims taking into account the relative values of each community. However, we do have good measures in both our studies for the status of the victim in each case. Specifically, for the Charging and Sentencing Study, we have 13 variables for the victim's physical characteristic, 18 variables for the victim's socioeconomic status, reputation, and criminal record. Adjustment for these variables did not diminish the estimated statewide race-of-victim effect. [4]

On the question of omitted variables, we want to point out that our data were derived from records of Georgia's Board of Pardon and Paroles. These records embody the results of extensive independent investigations carried out by Parole Board officials shortly after each homicide conviction. A representative of the Parole Board testified in the McCleskey case that these investigations are particularly thorough in homicide cases, and that investigators routinely examine court papers and all police records, and interview prosecutors, police officers, and witnesses. Given the tendency of such informants to justify their decisions, we would expect that any bias in their case reports would tend to emphasize the aggravating circumstances of the cases that received death sentences, which are commonly white-victim cases, and to emphasize the mitigating circumstances in

those that receive lesser sentences and which are more likely to have black victims. To the extent that this reporting bias exists, it would minimize the magnitude of the race-of-victim effects which we estimated in our analyses. Finally, on the question of omitted variables, we note that the experts for the State of Georgia in the McCleskey case emphasized and argued earnestly that the omitted variables biased the race-of-victim estimated. However, in spite of 1,000 hours of work on the case, the state's experts were unable to suggest any variables which could plausibly explain away the race-of-victim effect because they were correlated with both the race of victim and the death-sentencing outcomes.

Professor Heckman's claim that the race-of-victim effect is "much weaker in urban areas" than in rural areas appears to be based upon table 65 of our 1983 report, which presented unadjusted death-sentencing rates for urban and rural areas. The race-of-victim disparities in that tabulation after adjustment for only the race of defendant show a 10 percentage point disparity in urban areas versus a 25 percentage point disparity in rural areas. However, when adjustment is made for legitimate background factors, the race-of-victim disparities in urban and rural areas are quite similar. Specifically, after simultaneous adjustment for more than 230 nonracial background factors, separate linear regression analyses estimated a 7-percentage point race-of-victim disparity in urban areas ($p = .13$) versus an 8-percentage point race-of-victim disparity ($p = .05$) in rural areas.[5]

Subsequent logistic multiple regression analyses of the data from the Procedural Reform Study using (urban/rural x race of victim) interaction terms show identical race-of-victim effects in urban and rural areas after adjustment for the core model of aggravating and mitigating circumstances; specifically, the logistic regression coefficients for both urban and rural places was $b = 3.2$ ($p < .01$).[6]

Professor Heckman complains that there is no "clear understanding of how the law is supposed to operate if it is nondiscriminatory" and that "missing from all these studies is a suitable benchmark for measuring fairness." The purpose of our 1983 report was not to discuss the legal standards used to evaluate the fairness of a capital sentencing system. Our report was a technical analysis of the empirical results. The discussion of the standards to be applied by the court was included in the briefs filed by the lawyers. Nevertheless, the entire Georgia research project was informed by our understanding

of the substantial jurisprudence which has developed in the last 20 years, specifying the standards for interpreting statistical evidence of classwide purposeful discrimination in a large discretionary decision-making process.[7] In a nutshell, a prima facie case of classwide intentional discrimination normally rests upon a multivariate analysis sufficient to support an inference that race or sex are influencing a significant number of decisions in the process. There is no legal requirement that the racial discrimination be uniformly distributed throughout the decision-making system under challenge.

As for the degree of predictability that is required to constitute a "fair" capital-sentencing system, the law is less clear. However, we have in the last decade developed a series of measures, based upon the teaching of Furman v. Georgia and Gregg v. Georgia, which estimate the degree to which a given death sentence can be meaningfully distinguished from other sentences that receive lesser punishments. However, the Supreme Court's decision in McCleskey v. Kemp (1987) has essentially overruled the requirement of basic rationality and consistency in death sentencing with respect to the objective factors of the cases.[8]

As for the relevance of the statistical proof to McCleskey's case, we note that it was offered in support of three legal claims, and that in the view of a minority of the justices in McCleskey each claim was cognizable and relevant under the substantive law. First, McCleskey relied upon a claim of classwide disparate treatment under the Equal Protection Clause protection clause of the Fourteenth Amendment. Justice Blackmun, in his dissent, found the evidence not only relevant but also sufficiently strong to make out a prima facie case. McCleskey's evidence of classwide discrimination was also deemed sufficient by Justice Brennan to support a claim of arbitrariness and capriciousness under the Cruel and Unusual Punishment provision of the Eighth Amendment, because it constituted substantial evidence that the race of the victim was influencing a substantial number of death-sentencing decisions and, as a result, created a significant risk that McCleskey's death sentence was arbitrarily imposed.

McCleskey's third claim was that his death sentence was the product of racial discrimination. To support this argument, McCleskey offered our analyses of the disposition of over 629 cases in Fulton County's criminal justice system. Because only 10 of these cases resulted in a death sentence, it was not possible to conduct a multivari-

ate analysis of the impact of race of victim on those decisions. However, there was sufficient sample size to study the decisions leading up to and including the prosecutorial decision to advance cases to a penalty trial after a capital murder verdict was returned by a jury. Race-of-victim effects were particularly strong in the plea bargaining decisions and in the post-trial determination of whether a case should be advanced to a penalty trial. The combination of this quantitative analysis, the presence of only one black juror on McCleskey's jury, and evidence that there were no guidelines or system of regular consultation or oversight to guide the exercise of discretion in the Fulton County district attorney's office persuaded both Justices Brennan and Blackmun that McCleskey had established by a preponderance of the evidence that his death sentence was a product of race-of-victim discrimination.

Finally, we agree with Professor Heckman that the correlation due to common judges or prosecutors in different cases may slightly bias the statistical significance of the race-of-victim disparity, but it does not bias the estimated size of that disparity. Further, we agree with Heckman that "evidence of no jurisdictional effects would bolster the conclusion of these studies." As indicated above, we have presented just such evidence and the race-of-victim effects in both the PRS and the CSS remain strong after adjustment for the identity of the circuit in which the case was processed.

Finally, we have reported nonparametric analyses which again confirm the finding of the significant race-of-victim effect.[9]

Notes

1. D. Baldus, G. Woodworth & C. Pulaski, Discrimination and Arbitrariness in Georgia's Charging and Sentencing System: A Preliminary Report (1983) (unpublished) (hereinafter cited as Baldus, Woodworth & Pulaski '83).

2. Id. at 100-04.

3. D. Baldus, G. Woodworth & C. Pulaski, Jr., EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS, ch. 6 & App. A (1987) (forthcoming) (hereinafter cited as "Equal Justice").

4. Baldus, Woodworth & Pulaski, '83, supra note 1 pp. 8-9 of Table DD.

5. Id. at Table 65.

6. Equal Justice, supra note 3 at Table 6-25.

7. See D. Baldus & J. Cole, Statistical Proof of Discrimination (1980); Finkelstein, The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases 80 COLUM. L. REV. 736 (1980).

8. McCleskey v. Kemp, 107 S. Ct 1756 (1987).

9. Baldus, Woodworth & Pulaski, Arbitrariness and Discrimination in the Administration of the Death Penalty: A Challenge to State Supreme Courts, 15 STETSON L. REV. 133, 192-93 (1986) (cross-tabulations based on two nonquantitative case culpability rating systems and a regression-based index of case culpability).

The mechanics of random juror selection

Thomas J. Marx

Marx Social Science Research, Inc.

I. Introduction

In North America the designation of people eligible to serve on juries has been left to the discretion of court clerks or local officials. Over the last three decades, criminal defendants have proven that the resulting pools of people available for jury service (jury pools) often differ on one or more aspects from the universe from which they were drawn.

These legal challenges to the composition of jury pools have caused many jurisdictions to order that members of such pools be drawn at random. This paper considers three approaches for the random selection of jury pools: systematic sampling, random sampling with fixed probability and variable sample size, and random sampling with variable probability and fixed sample size. Each approach is evaluated by four criteria: freedom from statistical bias, variance of obtained sample size about the target sample size, ease of application to paper lists of the universe, and ease of application to such lists on computer tape.

The paper also examines the question of when to eliminate certain classes of individuals who must not or may not serve. The classes are made up of those who: have been recently in a jury pool of the jurisdiction, are statutorily barred from jury service, or may opt whether or not to join a jury pool.

II. How one gets called for jury duty under the "key man system"

Most of us who reside permanently at one address also reside permanently on a number of lists kept by governmental agencies. One list might be a census of all inhabitants of the locality. Invariably, a list of registered voters exists. In my state, Massachusetts, the list of all adult inhabitants is the sampling frame for selecting pools of potential jurors (jury pools) for the Superior Court, while the locality's voter registration list is the sampling frame for selecting jury pools for the Federal District Court.

Typically, for each court session of a month's duration, the clerk of the court or some other official determines how many people are needed in the jury pool. The clerk then allocates to each locality in the jurisdiction a share of the pool that is roughly--sometimes very roughly--proportional to the population of the locality. When a locality receives its quota of jurors from the clerk it must send more than the required number of adult residents a jury questionnaire. When the questionnaires have been completed under oath and returned, a

determination of whether the individual is eligible to serve as a juror is made.

Until recently, this phase of juror selection was almost always accomplished through what is known as the "Key Man System." The key man is often an official or employee of the community who selects people from the list to be issued jury questionnaires.

The law usually requires impartial selection of potential jurors to safeguard the rights of defendants to a trial by "an impartial jury" and to "due process of the law" under the Sixth and Fourteenth amendments to the U.S. Constitution. While the intention of key persons may be impartial selection, the result, as proven by statisticians in case after case in jurisdictions throughout the United States, has been biased selection in which some groups have been under-represented.

When a defendant, through his or her lawyer and statistician, has succeeded in showing an under-representation of a particular sex or race in a jury pool, the judicial response has ranged from doing nothing to ordering a new trial. Defendants have shown with lower rates of favorable judicial response underrepresentation of certain ethnic groups, national origins, religions, younger citizens, the better educated and professionals.

III. Random selection of jury pools: a fairer and safer way

A judge who finds that blacks have been under-represented in the pool from which a grand jury was drawn is in a dilemma. Doing nothing may seem to the jurist to make empty verbiage of the Sixth and Fourteenth Amendments or may be impossible in light of the jurisdiction's case law precedents.

Remedies, like requiring re-indictment or retrial from a jury drawn from a racially representative pool, threatens the jury selection pool itself. The existing pool can be made more representative by augmenting it with blacks or the existing pool can be drawn by a method that approximates racial neutrality. Ordering either of these alternatives is a strong action for a trial judge to take. A third alternative, choosing a sub-pool of potential jurors that is racially balanced from the original pool, makes the original pool more racially imbalanced and subject to the same challenge in future cases.

Faced with a set of unworkable alternatives in these situations and the prospect of recurring challenges to the jury pool's composition, judges have asked statisticians involved in

these cases whether there was any selection method that could reduce the number of challenges to the composition of jury pools. "Random selection," answered the statisticians, often adding that the cost of operating and administering the jury pool selection system need not rise.

The kernel of random selection is the replacement of the human selector--the key man or woman--by a table of random numbers for paper lists or random number generating software (random number generators) for computer lists. Instead of a person choosing people from a list by idiosyncratic criteria, both conscious and unconscious, the jury pool is assembled by lottery.

This brings me to the purpose of this paper. Some of you may be asked to help set up random selection systems in your states or provinces. I want to share with you my thoughts on handling the mechanics of the process. My qualifications are fifteen years as a statistician and computer professional and six years of being an expert statistical witness in litigation challenging the composition of jury pools.

IV. Allocation of jury pool shares to the localities

The first step in assembling a jury pool is to assign to each locality a share that it must contribute to the pool. Absent superior data collected by all local governments or by the state or provincial government, the most recently conducted national census can be used to fix the shares of each locality.

Locality shares may be assigned so that the proportion of the jury pool a locality must furnish is equal to the proportion of the jurisdiction's adult population the locality has. In other words, localities are the single stratum in proportional, stratified sampling. In addition to fairly sharing the service and administrative burdens among communities, proportional locality contributions may forestall jury composition challenges that assert geographical unrepresentativeness.

V. Ineligible persons

Once the share of a locality is set, the selection process takes place. Frequently the universe of eligible jurors is not the list from which jury pool members are sampled. In my state, for example, aliens, those who have been in a state court jury pool within the preceding two years, some categories of government employees, members of the bar, educators in public institutions, health care providers and convicted felons are classes of people excluded from serving on state court juries.

VI. Exempt persons

In addition to defining people ineligible to serve, statutes sometimes allow certain individuals to elect exemption from jury duty. In

Massachusetts, persons over seventy and parents with children under fifteen may exempt themselves from jury duty.

People who fall in this category must not be screened out (even though this is sometimes done); else they cannot exercise their statutory right of electing whether or not to accept jury service.

VII. Random selection

A. Definition of random selection under sampling without replacement. In classical statistics, random selection means that the probability of selecting any given individual on any draw is constant. This implies sampling with replacement, for if an individual is removed from the universe after selection, that individual's probability of selection will henceforth be zero. Moreover, if the sample size is fixed, then on each successive draw the selection probability of individuals not yet chosen would have to increase.

Sampling with replacement in this context seems to me meaningless unless we require jurors to perform double duty during their tours of service. Jury duty is therefore an event that calls for sampling without replacement.

I am going to define random selection without replacement to mean that each individual has a constant probability of selection on a single draw whose result is either sending that individual a jury questionnaire or not. This definition can only be maintained if the sample size is not fixed in advance. If the sample size is fixed, then selection probability cannot be the same on each draw.

B. Production of random numbers. Most random numbers produced today, including those in tables of random numbers, emanate from the computer. These computer-generated digits are not random at all, but are utterly predictable. The digits result from mathematical algorithms. The digits of each "random" number become the seed of its successor which is a mathematical function of its seed. Eventually, after millions of digits, the cycle of digits repeats.

In what sense then are these "pseudo-random numbers," as they are called, random? They are random only in the sense that they pass tests human beings have devised to evaluate randomness. If the auto-correlation of digits at all possible lags is within sampling limits of zero and runs tests and uniform distribution of digit tests are statistically insignificant, one only knows that three tests of randomness have been passed. A fourth test might be failed since multivariable relationships may take on an infinity of forms. Do not put your complete trust in pseudo-random numbers from a computer. Be especially aware of tests that appraise the "randomness" of the output from your particular random number generator.

C. Oversampling. Between the mailing of questionnaires and the appointments to a jury pool, many people are lost. This attrition arises from a variety of sources: undelivered mail, nonresponse, bars to service, elective exemption and excusals. Suppose that historically the fraction of individuals who fall into any of these categories is f . Then $m=n/f$ individuals should be sampled to yield, after attrition, a jury pool of about n .

D. The variance of \hat{n} . Under my definition of sampling without replacement, in (A), the actual sample size will be obtained in two stages that may be modeled as

$$\hat{n} = pN + fm,$$

where f is used to estimate the present fractional jury pool yield, $p = m/N$ is the probability of selection and N is the universe size. Although p is a function of f , I believe their sample values will be independently determined. The variance of \hat{n} is then

$$\widehat{\text{var}}(\hat{n}-n) = p(1-p)N + f(1-f)m. \quad (1)$$

Substituting m/N for p in (1) yields

$$\widehat{\text{var}}(\hat{n}-n) = (1-m/N)m + f(1-f)m.$$

This tells us that whenever $(1-m/N)$ exceeds $f(1-f)$ the first term on the right will exceed the second term on the right. As m/N is typically less than .05 while f typically varies between .1 and .9, the first term will usually dominate the second.

VIII. Approaches to sampling

Before you is a list, on paper or computer tape, of the adult residents of a locality. You have also a list of those people from the community who have served in the state jury pool within the last two years. You are asked to choose randomly n adult citizens for the forthcoming session of criminal court in your county or parish.

A. Systematic sampling.

1. Theory. If you are not a statistical purist, systematic sampling appears to be a simple approach that allows you to specify the sample size in advance.

Given m and having counted N , the universe size, there is a constant, k , that satisfies $km=N$. Therefore,

$$k = \frac{N}{m} = \frac{1}{p} \quad (2)$$

where p is the probability of selection. This formula tells us that selection of every k th individual on the list of N individuals will yield the required sample.

As there are k possible sequences of m individuals, one sequence must be chosen. From a table of random numbers, randomly choose a number from 1 to k . That number, call it i , tells you that individual i will be the first person chosen. You will select individuals $i, i+k, i+2k, \dots, i+(m-1)k$ for your sample size of m .

One view of the sampling procedure is that each of k mutually exclusive and exhaustive samples had a $p=1/k$ chance of being drawn. The draw consisted of randomly choosing one of the k samples. The variance of \hat{n} under systematic sampling, from (1), is $f(1-f)m$. The first term drops out because pN (i.e. m) is fixed, apart from a negligible fluctuation that results because N/m usually yields a fractional quotient.

Two data properties that affect the bias and precision (inverse of the variance of the sample mean about the universe mean) of systematic sampling are periodic variation and autocorrelation related to the list order (1). Bias is the more serious problem in jury pool selection. For example, the same last name, unless it is very common, will not be included twice in the same jury questionnaire mailing because k will always be large (usually over 100). Since relatives more often than unrelated individuals have the same last name, a jury pool derived from systematic sampling will have fewer members who have relatives on it than a randomly selected jury pool.

2. Practice. For communities with paper lists, one would first count N , then compute $k=N/m$. Choose a starting individual between 1 and k and take every k th person thereafter.

For communities with lists on tapes, generate a random number greater than or equal to zero and less than unity. Multiply the number by k , add one, and truncate to the nearest integer to get i , the first individual. Modular arithmetic can select every k th person.

Assuming sequential reading of records, target records may be screened by

IF MOD(R-I,K)=0 THEN ACCEPT RECORD.

Here, R is the sequence number of the record just read.

Systematic sampling is the easiest method I know of to implement on paper lists. This is its outstanding, and perhaps dominating, attraction. As any method is easy to program for a scientific programmer, implementation ease on the computer is not an issue for any of the three sampling approaches.

I have two reservations about systematic sampling. One, bias, has already been discussed. The second is that since lists change slowly, the same k in successive years may result in

choosing runs of people who were tapped for the jury pool in the previous year. This will be true even if the starting individuals in the two years differ.

A way to handle this problem is to alternately range up or down from the name of any person that has served within the previous two years until the name of a person who hasn't served appears.

Only people with prior service within two years would be screened out without being sent a questionnaire. Those who are barred from service on some other ground will have to swear that they are prohibited from serving.

B. Random sampling for variable m and fixed p

1. Theory. This is simple random sampling without replacement and few approaches could be simpler conceptually. Set $p=m/N$ and sample away.

The variance of n , as shown in the derivation of (1), will be $p(1-p)N+f(1-f)m$. The first term will usually dominate the second as argued in VII.D.

2. Practice. For paper lists, multiply p by 1000 and round off to obtain a three digit random number. Subtract 1 from this result. Call the number A . Choose three consecutive digits at random from a random number table. If the random number table number is less than or equal to A , take the first person on the list; otherwise reject the first person. Proceed to the next three random digits and the next name and repeat the procedure. Continue in this fashion through the entire paper list.

For computer tape, the selection statement is

```
IF P <= RANDOM(0) ACCEPT RECORD.
```

The advantages of this method are its conceptual simplicity and known freedom from sampling bias (apart from not replacing individuals).

In pseudo computer statements, the key instructions are

```
I=TRUNC(K*RANDOM(0)+1),
```

where 0 is a dummy argument to the random number generator.

There are two drawbacks. The first is that paper sampling is at least an order of magnitude slower than paper sampling under systematic sampling. Each individual must be tested against the random number table rather than every k th individual being checked off the list. The more complex and tedious procedure will invite errors as well as fudging.

The second drawback is that the sample drawn will be much further from n than under systematic sampling because we must add to the variance of estimating n from fm the much larger variance of pN .

C. Random sampling for fixed m and variable p

1. Theory. The difference between this type of sampling and sampling for fixed p and variable m is that in the latter approach, sample size will vary about m due to p being a random variable. In this approach we can adjust p from trial to trial so that exactly m individuals are selected.

Let $p_0=m/N$ on the first draw. If the first person is chosen, $p_1=(m-1)/(N-1)$ on the second draw; otherwise, $p_1=m/(N-1)$ on the second draw. Continue through subsequent draws in this way. If the individual is selected, decrease both numerator and denominator by one.

The procedure ensures that although p will vary from trial to trial, it will always oscillate around m/N . This may be shown by noting that if the first individual is drawn, then $p=(m-1)/(N-1)<m/N$ for the second person. Similarly, if the first person is not drawn, then $p=m/(N-1)>m/N$ for the second person. The argument may be extended to show that the further away from p the probability on a particular trial is, the more likely it will be that the next trial will result in movement back toward p . As in systematic variance, the variance of \hat{n} will be $f(1-f)m$. The procedure has eliminated the first term in (1) from the variance.

2. Practice. To allow p to vary while fixing m is a formidable task with paper lists and random number tables. Without getting into detail, one must track the declines in both m and N through the tables as well as redrawing several times before being able to determine whether many of the individuals will be sent the jury questionnaire. An alternative is to work with a calculator and random number table. This makes life easier with random numbers, but opens up the procedure to computation errors. In my judgement, this sampling approach won't work on paper lists.

The task is practical on a computer. On any draw

```
IF M/N<=RANDOM(0) THEN DO;  
ACCEPT RECORD; M=M-1;  
END;  
N=N-1
```

determines whether to mail or not mail to each individual and makes appropriate adjustments to m and N . n will be estimated as closely as under systematic sampling.

II. Summary

In the Key Man System of jury pool selection a selector chooses potential jurors from a locality. This system usually produces underrepresentation of identifiable groups and has spawned numerous challenges to the composition of the jury under the Sixth and Fourteenth Amendments to the U.S. Constitution, statutes and case law.

To assure jury pools that mirror the eligible universe, jurisdictions have begun to select jury pool members by methods that approach random selection. Random selection may be regarded as proportional, stratified, random sampling with localities as the stratum.

Before choosing jury questionnaire recipients, all persons who are excluded because of previous jury duty within a prescribed period should be screened out. Everyone else, including those who may be barred from jury service, should be eligible to receive a jury questionnaire. Answered under oath, this questionnaire, when returned, may be used to determine eligibility.

If n individuals are wanted from a locality $m=n/f$ should be sampled 'f' will be the historic fraction who become jury pool members of the people sent jury questionnaires.

I consider three ways to approximate random sampling: systematic, sampling for variable m with fixed p , and sampling for fixed m with variable p . For communities working with paper lists, systematic sampling is by far the easiest, variable m with fixed p at least an order of magnitude harder, and fixed m with variable p unworkably hard. On the computer, all methods are simple for a scientific programmer.

Systematic sampling has the greatest danger of yielding a biased sample. The other two methods seem to be safe from bias, but the third method will produce a final sample closer to n than will the second method.

My recommendation is systematic sampling for paper lists and fixed m variable p sampling for computer tape. If you are more comfortable with consistent methods for paper lists and computer lists, the simplicity of systematic sampling outweighs the risk that samples will be biased on any aspect a judge would take action on.

Notes

- (1) Cochran, William G., Sampling Techniques, 3rd ed., Sections 8.8 and 8.9, pages 217-221, Wiley: New York, 1977.

Appendix A: Program

Law and Justice Statistics Workshop Las Vegas, Saturday and Sunday August 3-4, 1985

NOTE: All papers listed were presented.
Those not included in the proceedings volume were either not
submitted by the speaker or, in the case of
John Rolph's, contained findings that
were superseded by subsequent studies.

Saturday, August 3

Session I: Statistical tools for law and justice administration, 9:00 a.m. - 12:00
Organizer -- Thomas A. Henderson, Criminal Justice Statistics Assoc., Inc.

Introductory remarks

Alan E. Gelfand, University of Connecticut
George G. Woodworth, University of Iowa

Redesign of the National Crime Survey and the Uniform Crime Reporting Program
Steven R. Schlesinger, Director, Bureau of Justice Statistics

Morning break

Some observations on the development of objective measures to aid decision-making
in the administration of justice

Paul F. Kolmetz, Director, Virginia Statistical Analysis Center
Richard P. Kern, Virginia Statistical Analysis Center

Lunch break

Session II: Statistics and probability in forensic science 2:00 p.m. - 5:00 p.m.
Organizer -- Lily E. Christ, John Jay College of Criminal Justice, CUNY

Against all odds

Carla M. Noziglia, Director of Crime Laboratory, Las Vegas Metropolitan Police
Department

Evaluating associative forensic science evidence

Barry D. Gaudette, Chief Scientist-Hair and Fibre, Royal Canadian Mounted Police

Afternoon break

Bayes theorem in forensic science

Piet de Jong, University of British Columbia

On identification by probability

Russell V. Lenth, University of Iowa

Sunday, August 4

**Session III: Statistical studies of the law and justice system:
Criminal careers 9:00 a.m. - 10:30 a.m.**
Organizer -- Alan E. Gelfand, University of Connecticut

Quantification and modeling of criminal careers

John P. Lehoczky, Carnegie-Mellon University
Alfred Blumstein, Carnegie-Mellon University

Morning break

**Session III: Statistical studies of the law and justice system:
Racial discrimination capital sentencing 10:50 a.m. - 12:00**
Organizers -- George G. Woodworth, University of Iowa
Colin Loftin, University of Maryland

Recent studies of race and victim effects in capital sentencing
George G. Woodworth, University of Iowa

Racial discrimination and arbitrariness in capital punishment: A review of the
evidence
Raymond Paternoster, University of Maryland

Comments on the Baldus, Woodworth and Pulaski, Gross and Mauro, and Paternoster
studies on disparity in capital sentencing by race of victim
James J. Heckman, University of Chicago

Lunch break

Session IV: Jury System Representativeness 2:00 p.m. - 4:00 p.m.
Organizer -- John E. Rolph, The RAND Corporation
Discussant -- Joseph B. Kadane, Carnegie-Mellon University

The mechanics of random jury selection
Thomas J. Marx, Marx Social Science Research, Inc, Cambridge, MA

Assigning jurors to court locations: Representativeness vs other criteria
John E. Rolph, The RAND Corporation

Appendix B: Addresses of speakers and organizers

Speakers

Dr. David C. Baldus
Department of Statistics
University of Iowa
Iowa City, IA 52242

Dr. Alfred Blumstein
Urban Systems Institute
School of Urban and Public Affairs
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Piet de Jong
Faculty of Commerce
and Business Administration
University of British Columbia
Vancouver, BC V6T1YH
Canada

Dr. Barry D. Gaudette
Chief Scientist - Hair and Fiber
Royal Canadian Mounted Police
Central Forensic Laboratory
1200 Alta Vita Drive
Ottawa, Ontario K1G3M8
Canada

Dr. Steven Gottfredson
Department of Criminal Justice
Gladfelter Hall
Temple University
Philadelphia, PA 19122

Dr. Samuel R. Gross
School of Law
University of Michigan
Ann Arbor, MI 48109

Dr. James J. Heckman
University of Chicago
Social Science Building, Room 405
1126 E. 59th Street
Chicago, ILL 60637

Dr. Joseph B. Kadane
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Richard P. Kern
Statistical Analysis Center
Department of Criminal Justice
805 E. Broad Street
Richmond, VA 23219

Dr. Paul F. Kolmetz, Director
Statistical Analysis Center
Department of Criminal Justice
805 E. Broad Street
Richmond, VA 23219

Dr. John P. Lehoczky
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Russell V. Lenth
Department of Statistics
and Actuarial Science
101 MLH
University of Iowa
Iowa City, IA 52242

Dr. Thomas J. Marx
Marx Social Science Research, Inc.
196 Appleton Street
Cambridge, MA 02138

Dr. Carla M. Noziglia
Director, Crime Laboratory
Las Vegas Metropolitan
Police Department
2575 South Highland Drive
Las Vegas, NV 89101

Dr. Raymond Paternoster
Institute of Criminal Justice
and Criminology
University of Maryland
College Park, MD 20742

Dr. Charles A. Pulaski
College of Law
Arizona State University
Tempe, AZ 85287

Dr. Steven R. Schlesinger, Director
Bureau of Justice Statistics
U.S. Department of Justice
633 Indiana Avenue
Washington, DC 20531

Session Organizers

Dr. Alan E. Gelfand
Department of Statistics
U-120
University of Connecticut
Storrs, CT 06268

Dr. Lily E. Christ
Department of Mathematics
John Jay College
CUNY
444 W. 56th Street
New York, NY 10019

Dr. Thomas A. Henderson
Criminal Justice Statistics Assoc., Inc.
444 North Capitol Street, Suite 606
Washington, DC 20001

Dr. Colin Loftin
Institute of Criminal Justice
and Criminology
University of Maryland
College Park, MD 20742

Dr. John E. Rolph
The RAND Corporation
1700 Main Street
Santa Monica, CA 90406

Dr. George G. Woodworth
Department of Statistics
University of Iowa
Iowa City, IA 52242

The 1986 Directory of Automated Criminal Justice Information Systems

If your company's market researchers want to know which Federal, State, and local criminal justice agencies are fully automated and what computer equipment is currently used, they need a copy of the Bureau of Justice Statistics' *1986 Directory of Automated Criminal Justice Information Systems*. This one-of-a-kind index lists more than 1,000 computerized information systems being used by police, courts, corrections, and other criminal justice agencies across the United States.

Organized alphabetically by State, city, or county, the *Directory* is a reference guide to information systems that are operational or are being developed. Each entry lists the type of information system in place—whether it be police computer-aided dispatch or Prosecution Management Support System (PMSS). In addition, the *Directory* supplies information about the status of a system's applications and its statistical and communications capabilities, names hardware and software, and furnishes

key contact names, addresses, and telephone numbers of criminal justice agency administrators and data processing personnel with purchasing authority.

Never before have so many aspects of criminal justice database information systems been systematically compiled and reported. Five indexes help locate systems by jurisdiction, system name, system function, statistical topic, and central processing unit.

The *Directory*, prepared by SEARCH Group, Inc., for the U.S. Bureau of Justice Statistics, is a major step in the Bureau's program to provide a current reference for data processing and criminal justice planners who are developing new systems or who are enhancing existing ones.

The *1986 Directory of Automated Criminal Justice Information Systems* costs only \$20.

Ordering information

Yes! Please send me _____ copy(s) of the *1986 Directory of Automated Criminal Justice Information Systems* (NCJ 102260).

To speed the delivery of your *Directory*, have your credit card ready and dial toll-free 800-732-3277.

Please print:

Name _____

Organization _____

Address _____

City _____ State _____

ZIP _____ Telephone _____

Check this box if you want to be placed on the Bureau of Justice Statistics mailing list.

To order, fill out and return this form with your check or money order. PREPAYMENT IS REQUIRED. Mail to: Justice Statistics Clearinghouse/NCJRS
Department F-AEY
Box 6000
Rockville, MD 20850

Enclosed is a check or money order in the amount of \$_____. (Please make check payable to the Justice Statistics Clearinghouse/NCJRS.)

Charge \$_____ to my VISA MasterCard

Account number _____

Signature _____

Expiration date _____

Crime and Older Americans

Information Package

- Are older Americans more likely to be victims of crime than younger age groups?
- Are the elderly being arrested for certain crimes more frequently than in the past?
- Are offenders in crimes against the elderly more likely to be strangers or nonstrangers compared to other age groups?

A new information package available from the Justice Statistics Clearinghouse answers these and other questions about crime and the elderly. Drawing from national sources for crime statistics—including the BJS National Crime Survey, the FBI Uniform Crime Reports, and the BJS National Corrections Reporting Program—the 34-page package discusses the types of crimes in which older Americans are most likely to be victims and offenders, and the types of crime prevention they use.

As the elderly population has grown, so has concern about the effects of crime on this age group.

Population statistics indicate that older Americans are fast becoming a large segment of the total U.S. population. In 1985, Americans 60 years and older totaled 39.5 million—a 21-percent increase over the past 10 years.

This package also includes the names and addresses of associations and organizations that are sources of information about crime and older Americans and a list of further readings.

Crime and Older Americans costs only \$10.00.

Please send me _____ copies of the Information Package on Crime and Older Americans (NCJ 104569) at \$10.00 each.

Name: _____

Organization: _____

Address: _____

City, State, ZIP: _____

Telephone: _____

Please detach this form and mail it, with payment, to:
Justice Statistics Clearinghouse
Dept. F-AGK
Box 6000
Rockville, MD 20850

Method of payment

- Payment of \$ _____ enclosed
- Check payable to NCJRS
- Money order payable to NCJRS

Please bill my

- NCJRS deposit account

Credit card Visa MasterCard

_____ Exp. date: _____

Signature: _____

Bureau of Justice Statistics reports

(revised June 1988)

Call toll-free 800-732-3277 (local 301-251-5500) to order BJS reports, to be added to one of the BJS mailing lists, or to speak to a reference specialist in statistics at the Justice Statistics Clearinghouse, National Criminal Justice Reference Service, Box 6000, Rockville, MD 20850. Single copies of reports are free; use NCJ number to order. Postage and handling are charged for bulk orders of single reports. For single copies of multiple titles, up to 10 titles are free; 11-40 titles \$10; more than 40, \$20; libraries call for special rates.

Public-use tapes of BJS data and other criminal justice data are available from the Criminal Justice Archive and Information Network, P.O. Box 1248, Ann Arbor, MI 48106 (313-763-5010).

National Crime Survey

Criminal victimization in the U.S.:

1985 (final report), NCJ-104273, 5/87
1984 (final report), NCJ-100435, 5/86
1983 (final report), NCJ-96459, 10/85

BJS special reports:

Motor vehicle theft, NCJ-109978, 3/88
Elderly victims, NCJ-107676, 11/87
Violent crime trends, NCJ-107217, 11/87
Robbery victims, NCJ-104638, 4/87
Violent crime by strangers and nonstrangers, NCJ-103702, 1/87
Preventing domestic violence against women, NCJ-102037, 8/86
Crime prevention measures, NCJ-100438, 3/86

The use of weapons in committing crimes, NCJ-99643, 1/86

Reporting crimes to the police, NCJ-99432, 12/85

Locating city, suburban, and rural crime, NCJ-99535, 12/85

The risk of violent crime, NCJ-97119, 5/85

The economic cost of crime to victims, NCJ-93450, 4/84

Family violence, NCJ-93449, 4/84
Criminal victimization 1986, NCJ-106989, 10/87

BJS bulletins:

Households touched by crime, 1987, NCJ-111240, 5/88

Criminal victimization 1986, NCJ-106989, 10/87

Households touched by crime, 1986, NCJ-105289, 6/87

The crime of rape, NCJ-96777, 3/85

Household burglary, NCJ-96021, 1/85

Violent crime by strangers, NCJ-80829, 4/82

Crime and the elderly, NCJ-79614, 1/82

Measuring crime, NCJ-75710, 2/81

The seasonality of crime victimization, NCJ-111033, 6/88

Series crimes: Report of a field test (BJS technical report), NCJ-104615, 4/87

Crime and older Americans information package, NCJ-104569, \$10, 5/87

Lifetime likelihood of victimization, (BJS technical report), NCJ-104274, 3/87

Teenage victims, NCJ-103138, 12/86

Response to screening questions in the National Crime Survey (BJS technical report), NCJ-97624, 7/85

Victimization and fear of crime: World perspectives, NCJ-93872, 1/85

The National Crime Survey: Working papers, vol. I: Current and historical perspectives, NCJ-75374, 8/82

vol. II: Methodological studies, NCJ-90307, 12/84

Issues in the measurement of victimization, NCJ-74682, 10/81

The cost of negligence: Losses from preventable household burglaries, NCJ-53527, 12/79

Repeat victimization in 26 American cities, NCJ-55878, 8/79

An introduction to the National Crime Survey, NCJ-43732, 4/78

Local victim surveys: A review of the issues, NCJ-39973, 8/77

Corrections

BJS bulletins and special reports:

Prisoners in 1987, NCJ-110331, 4/88
Profile of State prison inmates, 1986, NCJ-109926, 1/88

Capital punishment 1986, NCJ-106483, 9/87

Imprisonment in four countries, NCJ-103967, 2/87

Population density in State prisons, NCJ-103204, 12/86

State and Federal prisoners, 1925-85, 102494, 11/86

Prison admissions and releases, 1983, NCJ-100582, 3/86

Examining recidivism, NCJ-96501, 2/85

Returning to prison, NCJ-95700, 11/84

Time served in prison, NCJ-93924, 6/84

Historical statistics on prisoners in State and Federal institutions, yearend 1925-86, NCJ-111098, 6/88

Correctional populations in the U.S., 1985, NCJ-103957, 2/88

1984 census of State adult correctional facilities, NCJ-105585, 7/87

Historical corrections statistics in the U.S., 1850-1984, NCJ-102529, 4/87

1979 survey of inmates of State correctional facilities and 1979 census of State correctional facilities:

BJS special reports:

The prevalence of imprisonment, NCJ-93657, 7/85

Career patterns in crime, NCJ-88672, 6/83

BJS bulletins:

Prisoners and drugs, NCJ-87575, 3/83

Prisoners and alcohol, NCJ-86223, 1/83

Prisons and prisoners, NCJ-80697, 2/82

Veterans in prison, NCJ-79232, 11/81

Census of jails and survey of jail inmates:

Drunk driving, NCJ-109945, 2/88

Jail inmates, 1986, NCJ-107123, 10/87

Jail inmates 1985, NCJ-105586, 7/87

The 1983 jail census (BJS bulletin), NCJ-95536, 11/84

Census of jails, 1978: Data for individual jails, vols. IV, Northeast, North Central, South, West, NCJ-72279-72282, 12/81

Profile of jail inmates, 1978, NCJ-65412, 2/81

Parole and probation

BJS bulletins:

Probation and parole 1986, NCJ-108012, 12/87

Probation and parole 1985, NCJ-103683, 1/87

Setting prison terms, NCJ-76218, 8/83

BJS special reports:

Time served in prison and on parole, 1984, NCJ-108544, 1/88

Recidivism of young parolees, NCJ-104916, 5/87

Parole in the U.S., 1980 and 1981, NCJ-87387, 3/86

Characteristics of persons entering parole during 1978 and 1979, NCJ-87243, 5/83

Characteristics of the parole population, 1978, NCJ-66479, 4/81

Children in custody

Public juvenile facilities, 1985 (bulletin), NCJ-102457, 10/86

1982-83 census of juvenile detention and correctional facilities, NCJ-101686, 9/86

Expenditure and employment

BJS bulletins:

Justice expenditure and employment: 1985, NCJ-104460, 3/87

1983, NCJ-101776, 7/86

1982, NCJ-98327, 8/85

Justice expenditure and employment in the U.S.: 1980 and 1981 extracts, NCJ-96007, 6/85

1971-79, NCJ-92596, 11/84

Courts

BJS bulletins:

State felony courts and felony laws, NCJ-106273, 8/87

The growth of appeals: 1973-83 trends, NCJ-96381, 2/85

Case filings in State courts 1983, NCJ-95111, 10/84

BJS special reports:

Felony case-processing time, NCJ-101985, 8/86

Felony sentencing in 18 local jurisdictions, NCJ-97681, 6/85

The prevalence of guilty pleas, NCJ-96018, 12/84

Sentencing practices in 13 States, NCJ-95399, 10/84

Criminal defense systems: A national survey, NCJ-94630, 8/84

Habeas corpus, NCJ-92948, 3/84

State court caseload statistics, 1977 and 1981, NCJ-87587, 2/83

Sentencing outcomes in 28 felony courts, NCJ-105743, 8/87

National criminal defense systems study, NCJ-94702, 10/86

The prosecution of felony arrests:

1982, NCJ-106990, 5/88

1981, NCJ-101380, 9/86, \$7.60

1980, NCJ-97684, 10/85

1979, NCJ-86482, 5/84

Felony laws of the 50 States and the District of Columbia, 1986, NCJ-105066, 2/88, \$14.70

State court model statistical dictionary, Supplement, NCJ-98326, 9/85

1st edition, NCJ-62320, 9/80

State court organization 1980, NCJ-76711, 7/82

Computer crime

BJS special reports:

Electronic fund transfer fraud, NCJ-96666, 3/85

Electronic fund transfer and crime, NCJ-92650, 2/84

Electronic fund transfer systems fraud, NCJ-100461, 4/86

Computer security techniques, NCJ-84049, 9/82

Electronic fund transfer systems and crime, NCJ-83736, 9/82

Expert witness manual, NCJ-77927, 9/81, \$11.50

Criminal justice resource manual, NCJ-61550, 12/79

Privacy and security

Privacy and security of criminal history information: Compendium of State legislation: 1984 overview, NCJ-98077, 9/85

Criminal justice information policy: Automated fingerprint identification systems: Technology and policy issues, NCJ-104342, 4/87

Criminal justice "hot" files, NCJ-101850, 12/86

Data quality policies and procedures: Proceedings of a BJS/SEARCH conference, NCJ-101849, 12/86

Crime control and criminal records (BJS special report), NCJ-99176, 10/85

State criminal records repositories (BJS technical report), NCJ-99017, 10/85

Data quality of criminal history records, NCJ-98073, 10/85

Intelligence and investigative records, NCJ-95787, 4/85

Victim/witness legislation: An overview, NCJ-94365, 12/84

Information policy and crime control strategies (SEARCH/BJS conference), NCJ-93926, 10/84

Research access to criminal justice data, NCJ-84154, 2/83

Privacy and Juvenile justice records, NCJ-84152, 1/83

Federal justice statistics

The Federal civil justice system (BJS bulletin), NCJ-104769, 7/87

Employer perceptions of workplace crime, NCJ-101851, 7/87

Federal offenses and offenders

BJS special reports:

Pretrial release and detention: The Bail Reform Act of 1984, NCJ-109929, 2/88

White-collar crime, NCJ-106876, 9/87

Pretrial release and misconduct, NCJ-96132, 1/85

BJS bulletins:

Bank robbery, NCJ-94463, 8/84

Federal drug law violators, NCJ-92692, 2/84

Federal justice statistics, NCJ-80814, 3/82

General

BJS bulletins and special reports:

International crime rates, NCJ-110776, 5/88

Tracking offenders, 1984, NCJ-109686, 1/88

BJS telephone contacts '87, NCJ-102909, 12/86

Tracking offenders: White-collar crime, NCJ-102867, 11/86

Police employment and expenditure, NCJ-100117, 2/86

Tracking offenders: The child victim, NCJ-95785, 12/84

Tracking offenders, NCJ-91572, 11/83

Victim and witness assistance: New State laws and the system's response, NCJ-87934, 5/83

Report to the Nation on crime and justice, second edition, NCJ-105506, 6/88

BJS data report, 1987, NCJ-110643, 5/88

BJS annual report, fiscal 1987, NCJ-109928, 4/88

Data center & clearinghouse for drugs & crime (brochure), BC-000092, 2/88

Drugs and crime: A guide to BJS data, NCJ-109956, 2/88

Sourcebook of criminal justice statistics, 1986, NCJ-105287, 9/87

1986 directory of automated criminal justice information systems, NCJ-102260, 1/87, \$20

Publications of BJS, 1971-84: A topical bibliography, TB030012, 10/86, \$17.50

BJS publications: Selected library in microfiche, 1971-84, PR030012, 10/86, \$203 domestic

National survey of crime severity, NCJ-96017, 10/85

Criminal victimization of District of Columbia residents and Capitol Hill employees, 1982-83, NCJ-97982; Summary, NCJ-98567, 9/85

DC household victimization survey data base: Study implementation, NCJ-98595, \$7.60

Documentation, NCJ-98596, \$6.40

User manual, NCJ-98597, \$8.20

How to gain access to BJS data (brochure), BC-000022, 9/84

How to gain access to BJS data (brochure), BC-000022, 9/84

BJS maintains the following mailing lists:

- Drugs and crime data (new)
- White-collar crime (new)
- National Crime Survey (annual)
- Corrections (annual)
- Juvenile corrections (annual)
- Courts (annual)
- Privacy and security of criminal history information and information policy
- Federal statistics (annual)
- BJS bulletins and special reports (approximately twice a month)
- Sourcebook of Criminal Justice Statistics (annual)

To be added to these lists, write to: Justice Statistics Clearinghouse/NCJRS, Box 6000, Rockville, MD 20850.

See order form on last page

To be added to any BJS mailing list, please copy or cut out this page, fill in, fold, stamp, and mail to the Justice Statistics Clearinghouse/NCJRS.

You will receive an annual renewal card. If you do not return it, we must drop you from the mailing list.

To order copies of recent BJS reports, check here and circle items you want to receive on other side of this sheet.

Name:

Title:

Organization:

Street or box:

City, State, Zip:

Daytime phone number: ()

Criminal justice interest:

Put your organization and title here if you used home address above:

Please put me on the mailing list for—

- Justice expenditure and employment reports**—annual spending and staffing by Federal/State/local governments and by function (police, courts, etc.)
- White-collar crime**—data on the processing of Federal white-collar crime cases
- Privacy and security of criminal history information and information policy**—new legislation; maintaining and releasing intelligence and investigative records; data quality issues
- Federal statistics**—data describing Federal case processing, from investigation through prosecution, adjudication, and corrections

New!

- Juvenile corrections reports**—juveniles in custody in public and private detention and correctional facilities
- Drugs and crime data**—sentencing and time served by drug offenders, drug use at time of crime by jail inmates and State prisoners, and other quality data on drugs, crime, and law enforcement
- BJS bulletins and special reports**—timely reports of the most current justice data
- Prosecution and adjudication in State courts**—case processing from prosecution through court disposition, State felony laws, felony sentencing, criminal defense

- Corrections reports**—results of sample surveys and censuses of jails, prisons, parole, probation, and other corrections data
- National Crime Survey reports**—the only regular national survey of crime victims
- Sourcebook of Criminal Justice Statistics** (annual)—broad-based data from 150 + sources (400 + tables, 100 + figures, index)
- Send me a form to sign up for **NIJ Reports** (issued free 6 times a year), which abstracts both private and government criminal justice publications and lists conferences and training sessions in the field.

FOLD, SEAL WITH TAPE, AND STAMP

U.S. Department of Justice
Bureau of Justice Statistics
Washington, D.C. 20531

Place
1st-class
stamp
here

Justice Statistics Clearinghouse/NCJRS
U.S. Department of Justice
User Services Department 2
Box 6000
Rockville, MD 20850

Drugs & Crime Data

Data Center &
Clearinghouse for
Drugs & Crime



Illicit drugs— Cultivation to consequences

The worldwide drug business

Cultivation & production
Foreign
Domestic

Distribution
Export
Transshipment
Import into U.S.

Finance
Money laundering
Profits

The fight against drugs

Enforcement
Border interdiction
Investigation
Seizure & forfeiture
Prosecution

Consumption reduction
Prevention
Education
Treatment

Consequences of drug use

Abuse
Addiction
Overdose
Death

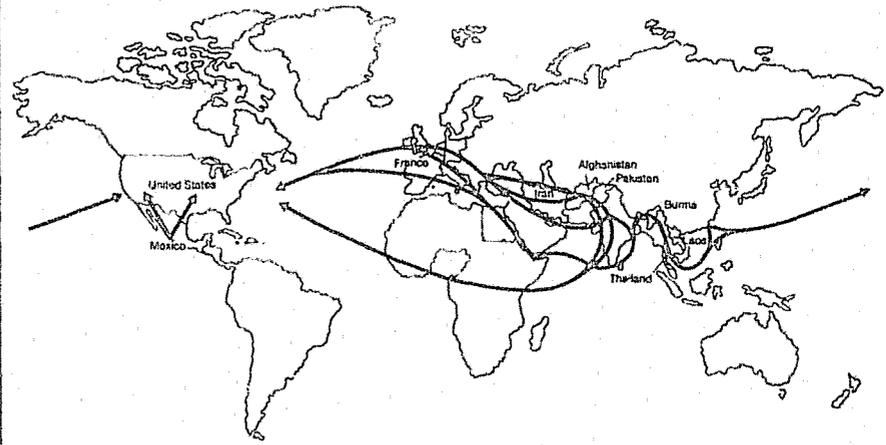
Crime
While on drugs
For drug money
Trafficking

Impact on justice system

Social disruption

The Data Center & Clearinghouse for Drugs & Crime is funded by the Bureau of Justice Assistance and directed by the Bureau of Justice Statistics of the U.S. Department of Justice.

Major heroin smuggling routes into the United States



DEA Quarterly Intelligence Trends

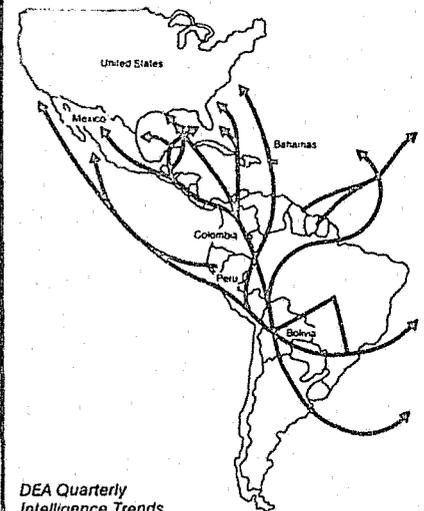
One free phone call can give you access to a growing data base on drugs & crime

The new Data Center & Clearinghouse for Drugs & Crime is managed by the Bureau of Justice Statistics. To serve you, the center will —

- Respond to your requests for drugs and crime data.
- Let you know about new drugs and crime data reports.
- Send you reports on drugs and crime.
- Conduct special bibliographic searches for you on specific drugs and crime topics.
- Refer you to data on epidemiology, prevention, and treatment of substance abuse at the National Clearinghouse for Alcohol and Drug Information of the Alcohol, Drug Abuse, and Mental Health Administration.
- Publish special reports on subjects such as assets forfeiture and seizure, economic costs of drug-related crime, drugs and violence, drug laws of the 50 States, drug abuse and corrections, and innovative law enforcement reactions to drugs and crime.

- Prepare a comprehensive, concise report that will bring together a rich array of data to trace and quantify the full flow of illicit drugs from cultivation to consequences.

Major cocaine smuggling routes into the United States



DEA Quarterly
Intelligence Trends

Call now and speak to a specialist in drugs & crime statistics:

1-800-666-3332

Or write to the Data Center & Clearinghouse for Drugs & Crime
1600 Research Boulevard
Rockville, MD 20850