

51054

Q-315-03158-



National Library of Canada

Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR Darryl Blair Plecas

TITLE OF THESIS/TITRE DE LA THÈSE The Perpetuation of Error in Criminology: A Case Study in Misreporting and the Failure of Organized Scepticism

UNIVERSITY/UNIVERSITÉ Simon Fraser University

DEGREE FOR WHICH THESIS WAS PRESENTED/ GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE M.A. (Criminology)

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE GRADE 1981

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE F. Douglas Cousineau

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE August 20, 1980 SIGNED/SIGNÉ [Signature]

PERMANENT ADDRESS/RÉSIDENCE FIXÉE [Redacted]



National Library of Canada
Collections Development Branch

Bibliothèque nationale du Canada
Direction du développement des collections

Canadian Theses on
Microfiche Service

Service des thèses canadiennes
sur microfiche

NOTICE

AVIS

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

THE PERPETUATION OF ERROR IN CRIMINOLOGY: A CASE STUDY IN
MISREPORTING AND THE FAILURE OF ORGANIZED SCEPTICISM

by

Darryl Blair Plecas

B.A. Simon Fraser University 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS (CRIMINOLOGY)

in the Department

of

Criminology

©

Darryl Blair Plecas 1980

SIMON FRASER UNIVERSITY

August, 1980

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without permission of the author.

APPROVAL

Name: Darryl Blair Plecas

Degree: Master of Arts (Criminology)

Title of Thesis: The Perpetuation of Error in Criminology:
A Case Study in Misreporting and the Failure
of Organized Scepticism.

Examining Committee:

Chairperson: Paul J. Brantingham

Douglas Cousineau
Senior Supervisor

Raymond Corrado

Alfred Keltner

Simon Verdun-Jones
External Examiner
Chairperson and Associate Professor,
Department of Criminology
Simon Fraser University

Date Approved: August 15, 1980

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

The Perpetuation of Error in Criminology: A Case Study in Misreporting
and the Failure of Organized Scepticism

Author: _____

(signature)

Darryl Blair Plecas

(name)

August 20, 1980

(date)

ABSTRACT

Confidence in the notion of science as a self-correcting enterprise is weakened when we encounter the perpetuation of significant error by influential social scientists. This thesis is a case study which shows that such error results from both the failure of social scientists to report information correctly and the failure of the scientific norm of "organized scepticism".

A methodological critique of the Special Intensive Parole Unit study (SIPU) reveals that, despite having been importantly and widely cited, the study is methodologically flawed. Serious methodological problems include inadequate definition of the independent variable - intensive parole supervision; a failure to adequately measure and report the dependent variable - recidivism; a failure to randomly assign parolees as intended; and a lack of controls to ensure only experimentals received treatment. A review of fifty-five citations to the SIPU study shows only three instances of attention to these flaws, and only one instance in which the study has been interpreted as methodologically unsound.

An examination of seventeen citations which cite each of the four phases of the SIPU study shows that only once have the findings been reported correctly. Also disclosed are instances of incorrect information being reported about the study's

design .

This thesis concludes by recommending the consideration of safeguards which would provide more reliable procedures for self-correction in science .

ACKNOWLEDGEMENT

I would like to thank the following for their assistance in my work on this thesis: my fellow students, and in particular Maureen Donovan, David Horne, Ronald Rea, Russell Smandych, and John Winterdyk who always showed interest and a willingness to comment; my friends, Ken Fernstrom and Bob Ward for their suggestions concerning editing matters; my professors in graduate course work, especially Dr. Patricia Brantingham, Dr. Paul Maxim and Professor Leslie Wilkins who provided continuous encouragement and numerous insights; and Aileen Sams, Graduate Program secretary, whose patience and kindness was more than one could ever hope for from a typist. I am especially grateful to my committee members, Dr. Raymond Corrado and Dr. Alfred Keltner for their critical comments and many helpful suggestions.

Finally, my senior supervisor, Dr. Douglas Cousineau, offered literally hundreds of suggestions concerning the direction and organization of my work, and to him I owe my greatest debt.

DEDICATION

To my wife, Joanne

TABLE OF CONTENTS

APPROVALii

ABSTRACTiii

ACKNOWLEDGEMENTv

DEDICATIONvi

LIST OF TABLESix

INTRODUCTIONx

I. THE PERPETUATION OF ERROR IN CRIMINOLOGY AND THE SIPU STUDY1

II. A CLASSIC STUDY REVISITED8

 INTRODUCTION8

 FIVE "INITIAL" CRITERIA9

 THE SIPU STUDY: PHASE I TO IV12

 THE SIPU STUDY AND CRITERIA #1 AND #2 (ADEQUATE DEFINITION AND MEASUREMENT)15

 THE SIPU STUDY AND CRITERION #3 (ADEQUATE SELECTION TECHNIQUES)21

 THE SIPU STUDY AND CRITERION #4 (ADEQUATE CONTROL OVER INDEPENDENT VARIABLE)27

 THE SIPU STUDY AND CRITERION #5 (ADEQUATE RESEARCH REPORT)29

 SUMMARY AND CONCLUSIONS30

III. MISREPORTING35

 INTRODUCTION35

 FINDINGS IN BRIEF37

 SELECTING A SAMPLE38

 FINDINGS AS REPORTED IN THE ORIGINAL REPORT46

DIFFERENT VERSIONS	49
MULTIPLE CONCLUSIONS AND A POOR PRESENTATION	51
SECONDARY SOURCES	52
INAPPROPRIATE LANGUAGE AND SELECTIVE CRITICISM	54
INADEQUATE DESCRIPTIONS, INACCURATE BASIC INFORMATION, INCOMPLETE REPORTS	58
CONCLUSIONS	61
IV. SUMMARY AND CONCLUSION: PREVENTING ERROR	63
INTRODUCTION	63
ENTRENCHED ERROR	64
TOWARDS SAFEGUARDS	67
APPENDIX A	71
APPENDIX B	73
APPENDIX C	75
APPENDIX D	77
BIBLIOGRAPHY	78

LIST OF TABLES

1	NUMBER AND NAMES OF CITATIONS TO SIPU RESEARCH.....	9
2	CITATIONS EXCLUDED FROM THE ANALYSIS.....	46
3	INCLUDED CITATIONS.....	49
4	REPORTS ON SIPU FINDINGS.....	55
5	CORRECT INCOMPLETE CITATIONS.....	65

INTRODUCTION

In social science there is an obvious need to guard against the perpetuation of error. In theory this is done through the process of "self-correction" (Campbell and Stanley, 1963:4-5; Kerlinger, 1973:6; Greer, 1969:7-8). Supposedly "self-correction" takes place because social scientists, being versed in established methodological standards, are guided by the scientific norm of "organized scepticism" (Merton, 1942:26;¹ Storer, 1966:78-79, 87-88). According to Storer this norm:

...is directive, embodying the principle that each scientist should be held individually responsible for making sure that previous research on which he bases his work is valid. (Storer, 1966:78).

Storer adds that:

The scientist is obligated also by the norm to make public his criticisms of the work of others when he believes it to be in error....It follows that no scientist's contribution to knowledge can be accepted without careful scrutiny, and that the scientist must doubt his own findings as well as those of others. He must hold himself entirely responsible for the goodness of his work. (Storer, 1966:79)

Clearly this is an immense responsibility.

Barnes and Dolby (1970:3-25) point out, however, that evidence for the existence of the norm of "organized scepticism" (and other norms) is not available and in fact has only been

¹Also see R.K.Merton "Science and the Social Order," Chapter 12 in R.K. Merton, The Sociology of Science, University of Chicago Press, 1973, (pp. 264-266). Chapter is reprinted from Philosophy of Science 5 (1938:321-337).

"professed" by scientists to exist². In a different vein, Mitroff (1974:579-595) acknowledges the existence of norms, but attempts to show that social scientists adhere to a set of counternorms. "Organized dogmatism" provides the counternorm for "organized scepticism". In describing this counternorm he quotes Boguslaw (1968:59):

Each scientist should make certain that previous work by others on which he bases his work is sufficiently identified so that others can be held responsible for inadequacies while any possible credit accrues to oneself. (Mitroff, 1974:592)

Similarly, Weinstein (1979:639-652), following a discussion of fraud in science, concludes that specialization has weakened the norm of "organized scepticism" to the extent that "detailed examination" has almost been replaced by "criteria such as prestige...or shared assumptions".

It is not surprising debate exists concerning "organized scepticism" and how it contributes to the self-correction process as "organized scepticism" is not a set of rigorous standard procedures for assessing research but merely an

²While Barnes and Dolby provide no evidence, they argue that scientists are only selectively sceptical as "tenacious advocacy of theories by individual scientists generally brings reward within a scientific sub-group sufficient to outweigh sanctions from outside it. The scientist generally has a specific viewpoint which makes him sceptical of some results while uncritically accepting others; this is the norm...at best we can claim that scientists possess a distinctive pattern of scepticism". (1970:10-11)(emphasis added)

expectancy of the social science community³. Confidence in "organized scepticism" is weakened further when we encounter the perpetuation of significant errors by prolific and influential social scientists. Furthermore, when these errors are found to be the result of "misreporting", as well as the failure of social scientists to exercise the norm of "organized scepticism", the process appears to be less a scientific enterprise than one which compounds error.

This thesis examines the perpetuation of a number of errors concerning the classical and seminal Special Intensive Parole Unit (SIPU) study and illustrates both the failure of "organized scepticism" to serve as a method of self-correction and the failure of social scientists to report information correctly. A major error is the belief that the SIPU study is a methodologically adequate experiment and is among the best research ever conducted in corrections. Additional problems involve the contention that the findings of the study provide knowledge on deterrence, the differential effectiveness⁴ of

³Underlying this expectancy is the belief that science's reward system (i.e. professional recognition) serves as a control mechanism (Storer, 1966:19-27; Merton, 1957:635-659; Campbell, 1979:181-201).

⁴"Differential effectiveness" means the effectiveness of the intervention strategy was different for some types of offenders than for others, or the effectiveness was differentially affected by the presence of one or more variables. It is the former meaning to which social scientists refer, although there seems no reason why the latter should not be recognized.

parole supervision, and on the effect of employment on recidivism among parolees. Also in error is the claim that the study's findings provide proof of the inefficacy of parole supervision. The two most serious errors involve the beliefs that the SIPU study provides part of the justification for both significant changes in social policy in the administration of criminal justice and directives for future research in criminology.

The analysis in this thesis shows there are drastic discrepancies between the originally published version of the SIPU study and most of the contemporary descriptions and evaluations of it in social science and criminology literature. The original research document reveals methodological inadequacies of such magnitude that no valid conclusions or social policy implications can be made from any individual phase of the study or from the four phases together. Additionally, an analysis of citations to the study shows a lack of accurate reports on the study's findings as reported in the original research document.

The methodological inadequacies of the SIPU study, its widespread but essentially uncritical acceptance, and the frequency with which the original findings have been misrepresented

clearly show the need to develop a more reliable means for self-correction. This thesis is offered as an attempt to draw critical attention to this need. In the final chapter new safeguards are proposed.

I. THE PERPETUATION OF ERROR IN CRIMINOLOGY AND THE SIPU STUDY

In the social sciences, research on the perpetuation of error is scant. Reynolds (1966:85-88) traces and documents the reporting of an unsubstantiated claim (i.e. women are better at color discrimination than men) taken as fact in psychology while Lauer (1971:881-889) argues that certain assumptions (i.e. change is traumatic, not normal) about social change are "scientific legitimations" of fallacies. Mackie (1973:431-447) shows that a major concept (i.e. stereotypes) in sociology lacks the empirical basis sociologists claim it has, and Morrison and Henkel (1969:131-140) record the persistent and inappropriate use of statistical techniques despite the considerable dispute about them in reputable social science journals.

The perpetuation of error in criminology apparently has not been studied. To be sure, there is no shortage of recent examples of where errors have occurred. Ross and McKay (1978:279-295) note that criminologists will be led to "four erroneous conclusions" unless a reassessment is made of the research on correctional treatment techniques, while Cousineau and Verdun-Jones (1979:295-297) draw attention to flaws in studies on plea bargaining. Roesch and Corrado (1979:530-541) found that a major and widely supported study of rehabilitation and diversion which claimed to show evidence of the inefficacy

of such programs "either glossed over or ignored completely". most of the basic or standard criteria for program evaluation. Savitz, Turner & Dickman (1977:41-56) argue that criminologists have wrongly pointed to Lombroso as the founder of "scientific" criminology (Positivist School) when this credit should go to Gall. Their examination of Gall's work and references to it, reveals that the "complete devaluation of phrenology as being completely unscientific is never based on the author's considered evaluation of the published work, research, and data developed by Gallian phrenologists".(1977:42).

Factors contributing to the perpetuation of error include the ways literature reviews are conducted (Yin, Bingham and Heald, 1976:139-156); the commonly accepted practice of citing secondary sources rather than originally published research (Sjoberg and Nett, 1968:324-325); and the misreporting of original research (Carey, 1967:408-416). Headley and Taveggia (1977:108-116) recently analyzed the process whereby textbook authors create and perpetuate error. Their research of textbook knowledge discloses significant discrepancies between what is described in textbooks and what is reported in the original research reports.. Headley and Taveggia found that discrepancies were the consequence of erroneous citations and the omission of relevant and pertinent detail (1977:112-113). Most disquieting is, as Weinstein (1979:639-652) reminds us, the fact that error in science can be created through fraud. She concludes that

"much fraud is likely to occur without being discovered" and refers to Cyril Burt's widely cited work on the relationship between I.Q. and heredity as a case in point. She notes that Burt was accused of having "doctored" or invented his I.Q. test data to support his theory on intelligence. An analysis documenting the virtual absence of correct citations to the SIPU study, lends significant credence to the conclusions of these students of the norm of self-correction.

For various reasons the SIPU study seems to be an appropriate point at which to begin an examination of the perpetuation of error in criminology. First, the SIPU study has been cited often. Over more than two decades at least fifty different social scientists have cited the study (see Table I below) in fifty-five publications. The publications include textbooks, anthologies, government reports, and prestigious journals. Thus information about the SIPU study has been available to criminologists, and we would expect that at least fifty social scientists would have given the study and its findings some scrutiny.

TABLE I

NUMBER AND NAMES OF CITATIONS TO SIPU RESEARCH

YEAR	NUMBER OF CITATIONS	AUTHORS
1959	1	Barnes & Teeters*;
1960	3	Gibbons; Sutherland & Cressey*; Tappan*;
1961	1	Reckless*;
1962	1	Cavan*;
1963	0	
1964	2	Clegg*; Glaser;
1965	1	Conrad;
1966	1	Manual of Corr. Standards;
1967	3	Keve; Reckless*; Adams;
1968	3	Outerbridge; Carter & Wilkins; Gibbons*;
1969	5	O'Leary; Dressler*; Robinson; Glaser(a); Glaser(b);
1970	1	Outerbridge
1971	4	Adams & Veeters; Robinson & Smith; Public Health; Harlow, Weber & Wilkins;
1972	2	Waller; Logan;
1973	4	Ledain Commission; Reckless*; Klapnuts; Gibbons*;
1974	4	Martinson; Waller; Adams; O'Leary;
1975	5	Citizens Inquiry; Lipton, Martinson & Wilks; McCleary; Kassebaum, Ward, Wilner; Gottfredson;
1976	5	Duffee & Fitch*; Adams; Bean; Solomon; Burkhardt;
1977	4	Maltz & McCleary; Greenberg; Lerner; Carney;
1978	4	Sutherland & Cressey*; National Parole Service; Brill; Boruch, McSweeney, Soderstrom;
1979	1	Miller;
Total	55	

* - Textbooks

Second, the study is considered excellent. Glaser (1964:457) describes the study as "Probably the most extensive controlled experiment in American correctional history, and perhaps in world history". Robinson and Smith (1971:70) claim the study employed "random assignment of cases" while Borch, McSweeney, and Soderstrom (1978:661) writing in the Evaluation Quarterly cite one phase of the study (Phase III) as "evidence" of the feasibility and scope of randomized field experiments. Finally, Lipton, Martinson, and Wilks (1975:1-161) categorize the study as nearly a "pure" design in terms of internal validity. These authors claim the criteria of Campbell and Stanley (1963:4-5) guided them in their quest for internally valid studies (see Lipton et al., 1975:16). Of two hundred and thirty-one studies examined in a meta evaluation by Lipton et al., only twelve other studies received as high a rating as was assigned one phase (i.e. Phase III) of the SIPU study¹.

¹Having passed initial screening standards, studies were subsequently evaluated in terms of how far their research designs satisfied the authors' criteria of internal validity. This resulted in a scale ranging from 1, the very best kind of design, to 18, the very worst kind of design. All studies were placed on the scale, and assigned a number to indicate the author's judgment on the methodological adequacy (Lipton, et al., 1975:17). In addition to assigning each research to a numerical category, a letter grade of A or B was assigned to qualify the numerical ranking by signifying in a "gross" way the reservations of the authors in their interpretations of the findings. The criteria for the assignment of the letter grade were not specified.

The combination of criteria, and the scoring of research designs by number and letter were applied to some 231 studies. Of these, only 13 studies were placed in numerical category 1.

Third, the SIPU study has been used as a source of alleged knowledge for a broad range of important matters in criminology. Specifically, the SIPU study has been regarded as providing knowledge on the question of deterrence (e.g. Lerner, 1977:213); providing knowledge on the question of the effect of employment on recidivism among parolees (e.g. Sutherland and Cressey, 1966:644-656; 1978:640-651; Tappan, 1960:738); providing information about prediction techniques (i.e. Gottfredson, 1975:86-87); providing evidence for the efficacy of parole supervision (e.g. Conrad, 1965:295; Canadian Parole Service memo, 1978:1-5; Dressler, 1969:268-269); providing evidence for the inefficacy of parole supervision (e.g. Carney, 1977:106; Miller, 1979:256-257, 262; Outerbridge, 1968:378-387; O'Leary, 1974:937; Gibbons, 1968:529, 1973:541-542); providing evidence of the inefficacy of correctional treatment generally (e.g. Kassebaum, Ward and Wilner, 1971:308; Citizens Inquiry, 1975:69-70, 104; Outerbridge, 1970:189-200; Robinson and Smith, 1971:76-77; Martinson, 1974:46-47; Greenberg, 1977:135-136); and providing directions for new research (differential effectiveness of specific techniques) (e.g. Glaser,

¹(cont'd) Numerical category 2 was the research design score given to the greatest number of studies. Lipton, et al judged 140 studies to have achieved this rather high score. The SIPU studies were awarded the following scores:

SIPU I - 2B (see p.153)
SIPU II - 2B (see p.156)
SIPU III - 1B (see p.160)
SIPU IV - 2B (see p.161)

1969(b)11-46; Brill, 1978:229; Soloman, 1976:266; Harlow, Weber and Wilkins, 1971: 343; Klapmuts, 1973:419).

Finally, the SIPU study seems an appropriate basis for analysis because criminologists have used SIPU findings as part of the basis for recommending important changes in criminal justice policy. SIPU study findings have been cited as part of the justification for important policy changes concerning parole in particular, (e.g. Outerbridge, 1970:189-200) and the corrections system in general (e.g. Martinson, 1974:22-54; Outerbridge, 1970:189-200).

Given the clearly demonstrated impact of the SIPU study on criminological theory and because it has been used as the basis for justifying criminal justice policy, it is not unreasonable to expect that it would have met at least minimum methodological criteria. Unfortunately, as the next chapter shows, the SIPU study left a host of validity concerns unanswered and, equally important, many others in considerable doubt.

II. A CLASSIC STUDY REVISITED

INTRODUCTION

One would expect that if criminologists are guided by the norm of "organized scepticism", then major methodological flaws in the research cited (in this case the SIPU study) would be identified. Major flaws would be easily identifiable and based on methodological criteria the social science community would recognize as appropriate, attainable, and well within generally accepted "minimal" standards of scientific adequacy (e.g. see Bernstein, 1978:24-47).¹

The analysis presented in this chapter provides an assessment of the SIPU research against five criteria, four of which are methodological. One criterion, while not methodological, is quite clearly a criterion for scientific adequacy. A reading of the original report on the SIPU research reveals that none of the study's four phases came close to

¹ Bernstein's analysis of social scientist's conformity to technical norms purports that such conformity is variable among evaluative researchers. Unfortunately, Bernstein's report of the analysis does not provide sufficient information concerning the sample of evaluative researchers studied to permit the reader to draw a more definitive conclusion. Equally unfortunate, the analysis did not provide an account of which of the (six) methodological prescriptions listed shows the greatest variability.

meeting any one of these five criteria.

FIVE "INITIAL" CRITERIA

Clearly, the criteria used here represent only a partial list of the criteria generally considered important in employing an experimental design, intended to test the efficacy of an intervention strategy. In fact, the criteria to be used here would better be described as "initial" criteria for if an experiment failed to meet any one of them, there would be no need for a more rigorous assessment of construct, statistical, external, or internal validity². These criteria are:

1. Adequate Definitions . The definitions provided for the dependent and independent variables must be sufficiently operational so the components of each are identifiable. This criterion is a pre-requisite to an assessment of replicability.
2. Adequate Measurement . The measurement of the dependent and independent variables must be such that they are interpretable and offer at least a modest indication of

² For an excellent discussion on the assessment of "validity" in experimental and quasi-experimental designs see; Cook, Cook, and Mark, 1977 (103-139) "Randomized and Quasi-experimental Designs in Evaluative Research: an Introduction" in L. Butman (ed.) Evaluative Research Methods: a Basic Guide, Beverly Hills: Sage Publications.

reliability and validity³.

3. Adequate Selection Techniques . There must be evidence that subjects were assigned to control and experimental groups on a random basis, or assigned through the use of matching techniques. This criterion is a pre-requisite to an assurance that an observed effect or non-effect is not due to a difference in the kinds of persons between and within groups.
4. Adequate Control Over Independent Variable . There must be some evidence that the experimental group is subjected to the intervention strategy as defined, but that the control group is not. This criterion is a pre-requisite to an assessment of the effect of the independent variable.
5. Adequate Research Report . The research report must provide sufficient information to permit replicability, verification, and the discovery of possible error in the data and research design. This criterion makes it possible for independent observers to exercise the scientific norm of "organized scepticism"⁴.

³ A measurement is considered to possess "reliability" when it is consistently accurate under varying conditions. A measurement is considered to possess "validity" when what we intend to measure is measured and not something else.

⁴ McTavish et al (1977:37) citing McTavish and Anderson (1973) note that one "criterion" for scientific research calls for researchers to "gather and analyze relevant data in a manner such that each possible conclusion is potentially falsifiable".

The above criteria are essentially a restatement of the "exclusion" criteria used by Lipton, Martinson, and Wilks (1975:6-7) in their compendious evaluation of evaluation studies. Lipton et al established eleven exclusion criteria which they used to rule out studies which were so inadequate they did not permit or warrant further assessment of methodological adequacy (see Appendix A)⁵. The criteria used for the present analysis are also consistent (although less inclusive) with the seven criteria Logan (1972:378-389) cites as the minimal methodological requirements a study must meet before it merits any further consideration as to its scientific adequacy (see Appendix B).

Again it is emphasized that the criteria used here are only pre-requisites to an assessment of scientific adequacy. These criteria were applied to the four phases of the SIPU study. Each phase is described below.

⁵Martinson, however, later criticized the procedures used in the selection of studies included in the Lipton et al survey because he felt they were too rigorous, needlessly rejecting significant pieces of work because of methodological flaws. He is quoted as saying, "we must set aside methodological fanaticism, which is the hallmark of every field that loses sight of substantive concerns" and it was reported that he added that research paid for by the public should be expected to produce some new information even if not methodologically perfect. (Reported in the Criminal Justice Newsletter; 1978, vol.9, 24, pg.4.) (emphasis added)

THE SIPU STUDY: PHASE I TO IV

In July of 1953, the California Department of Corrections (Division of Adult Paroles) began experimenting with reduced parole caseloads as a way of decreasing the number of parole violators returned to prison. It was assumed that reduced caseloads would offer more intensive supervision to parolees than was available in the usual parole caseload. The predicted results of intensive supervision was a lower recidivism rate. Under the Special Intensive Parole Unit (SIPU) Program the authorities designed and implemented four experiments to test the effectiveness of reduced caseloads.⁶ In each experiment the size of the reduced caseload differed, as did the length of intensive supervision. These experiments were referred to as Phase I:Fifteen Man Caseload Study (1956), Phase II:Thirty Man

⁶The SIPU Program also included a separately conducted and reported study (sub-study of Phase IV) which was designed to test the hypothesis that parolees with high base expectancy scores should require less supervision than parolees with lower scores. This study was known as "The High Base Expectancy Study". See Joan Havel, Research Report #10, Research Division, Department of Corrections, State of California (14pp)(no date). The study will not be referred to further in the present analysis because it does not address the issue of caseload sizes, as do the other research reports on the SIPU research. Care was taken to ensure that the citations used in the present analysis were not referring to the sub-study as opposed to the larger Phase IV study. Of all the citations to the SIPU research only three made reference to Research Report #10. These citations are Glaser (1969b:45), Lerner (1977:223), and Adams (1976:254). Lerner (1977) and Adams (1976) cite the Parole Outcome Study (Phase IV) as well. In another writing Glaser also refers to a "Phase V" study, although he cites no reference for it, and apparently it has not been mentioned elsewhere by anyone. (see Glaser 1969a:310-312;1964:458).

Caseload Study (1958), Phase III (1962), and Phase IV: Parole Outcome Study (1965).

Phase I began in February, 1954 and ended in mid-1955 (SIPU I:6). Prior to the study the average caseload size for the California Adult Parole Authority was about ninety parolees. The plan of the researchers was to establish a pool of parolable inmates and then to randomly assign them to either the regular, ninety man caseloads, or to intensive, fifteen man caseloads. It was believed each parolee assigned to the fifteen man caseloads would receive at least two hours per week of "very close contact" (SIPU I:1). This kind of supervision was to be provided for three months, followed by assignment to the regular ninety man caseloads. It was intended that the randomization process would allow for the systematic comparison of parolee success rates between the regular and intensive case loads⁷.

Phase II took place from July, 1957 to June, 1959 (SIPU II:5). In this phase the size of the experimental caseloads was increased from fifteen to thirty and it was intended that parolees assigned to these caseloads would spend six rather than three months under intensive supervision before being transferred to the ninety man caseloads. Again, it was intended that the randomization process would allow for the systematic

⁷This phase of the study was also concerned with the test of the practicability of releasing men on parole for an average of three months in advance of the usual time without hazard to the public safety. (see SIPU I:3).

comparison of parolee success rates between the regular and intensive caseloads⁸.

Phase III took place from July, 1957 to June, 1959 (SIPU III:4). In this phase caseload sizes were set at thirty-five for experimentals and seventy-two for controls. An additional design change from the first two phases was the elimination of the requirement that parolees transfer from reduced to regular sized caseloads after a set period of time. In this phase it was expected that experimentals would remain on the reduced caseloads for the duration of the study. As in the first two phases, success rates between parolees on the regular and intensive caseloads were compared after a twelve month follow-up period⁹.

In Phase IV, which ran from July 1959 to January, 1962, caseload sizes were set at fifteen, thirty, and seventy (SIPU IV:ii). A comparison of the success rates of the parolees assigned to each of the three caseload sizes was made after a twelve month followup¹⁰.

⁸ As per footnote #7, this phase was also concerned with advanced release (SIPU II:1).

⁹ Phase III also included a twenty-four month follow-up period (see SIPU III:1, 10-12).

¹⁰ This phase of the study was also designed to test differences among parolees of different maturity levels (see SIPU IV: 1-46).

THE SIPU STUDY AND CRITERIA #1 AND #2 (ADEQUATE DEFINITION AND MEASUREMENT)

While each of the four phases was designed to assess the impact of intensive supervision of parolees, it is astounding to discover that this variable was never defined or adequately measured. The concept of intensive supervision is essentially a qualitative one, similar to the idea of the gravity of crime (Nettler, 1978:60). Unfortunately, the researchers regarded caseload size and the number of "contacts" between parolees and their supervisors as the indicator of this concept. While there is always a concept-indicator problem in research (Nagel, 1961:93-94; Hirschi and Selvin, 1967:177-198) the attempt to simply substitute the number of contacts as a measure of the intensity of supervision confuses the frequency of supervision with the quality and intensity of supervision. This is a methodologically and inferentially dubious practice (Sellin and Wolfgang, 1964:1-4; Sjoberg and Nett, 1968:267-271; Nettler, 1978:60)¹¹

If, however, one is prepared to overlook the concept-indicator problems and simply accept the study as an attempt to determine the efficacy of reduced caseload sizes,

¹¹Carter and Takagi (1967:34-46) and Carter and Wilkins (1976:391-401) have argued that it is useless to use "caseload size" as a measure of supervision without specifying the components of supervision. This criticism is also pointed out by Adams, Chandler, and Neithercut (1971:46) in their critique of the San Francisco project.

then one still confronts the fact that caseload size itself was so poorly measured as to render each phase of the study nugatory.

First, parole officers were permitted some flexibility in transferring subjects from the experimental groups. The Phase I design, for example, called for the experimental subjects to remain under intensive supervision for three months, yet due to a variety of problems some parolees were supervised for only two months while others were kept on for as long as six months, (SIPU I:6). The Phase II design called for experimentals to receive six months intensive supervision, yet parole agents were allowed to transfer subjects as early as three months or retain them for as long as nine months (SIPU II:6). We are told by the Phase III research document that "except for a few cases" SIPU parolees did not remain under intensive supervision for the intended two-year period, but that "the range was six to twenty-four months" (SIPU III:11)¹² With respect to Phase IV, we are not even told how long parolees remained on the smaller caseloads. Despite these variations in actual supervision time, in each of the four phases data are presented for all parolees

¹²The average is reported as "16.5 months" (see SIPU III;11). The researchers suggest that it "reasonable to assume that transfers occurred with the same relative frequency in the control group (see SIPU III;4). This, however, seems a rather bold assumption given that they do not provide any information on the number of parolees transferred earlier. They refer only to "unknown numbers". Information in this regard was also lacking in the first two phases. (see SIPU III;4).

in the experimentals as if they were all supervised for the same time period.

Second, and probably most important, caseload size was defined as and measured by the number of contacts between parolees and their supervisors. In the Phase I report we are not informed of the nature of these "contacts" (ie. telephone conversations or personal interviews) and no tables or data on the number of contacts are provided. We are told only that the number of contacts for the experimentals "tended to average between eight and nine contacts a month" and ranged between wide limits (SIPU I;6). Not even crude guesses about the average number of contacts for the controls are provided. Similarly, the reports on Phases III and IV fail to provide any information about the kind or number of contacts received by experimentals and controls.

It is only in Phase II that the researchers give serious attention to the matter of contacts. The researchers report that although the number of contacts in both experimental and control groups varied within wide limits, experimentals averaged twice as many contacts as controls (SIPU II:9). But this statement is misleading. In order to estimate the frequency of contacts for the parolees on the small and large caseloads the researchers selected two samples. The first sample consisted of thirty parolees from the small caseloads and thirty from the large caseloads, while the second sample consisted of twenty-four from

large caseloads and twenty-seven parolees from the small caseloads (the approximate number of subjects in this phase of the study is 1,063 experimentals and 2,065 controls). The researchers reported the findings from this sampling indicated "a significant number of cases in each category received the same number of contacts from the parole agents" (SIPU II:9). Indeed, a re-analysis of the combined samples by this observer shows, however, that the "significant number" amounts to virtually half (49.5%) of the subjects sampled. Unfortunately, all data are presented in terms of the average number of contacts received, so we do not know how many experimentals did not actually receive intensive supervision as intended, or how many controls did receive intensive supervision as unintended. We do know from the available data though, that there is clear lack of evidence that only experimentals received treatment and good reason to believe that some experimentals received little if any treatment.

The next concern relates to problems in the measurement of the dependent variable. Apparently, in each phase of the study the outcome measures were unreliable and inadequately reported.

In Phase I the measure consisted of a "major arrest" defined as "arrest followed by conviction and sentence of sixty days or more" (SIPU I;7). The report states that this measure included "a few cases of obvious felony offenses by parolees that were not prosecuted" (SIPU I;7). The researchers provide no

information on the criteria for determining "obvious felony offences", nor do they tell us who made this judgment or how many of these cases there were. In addition, "many" (an unknown number) parolees "were hospitalized or referred to psychiatric facilities in lieu of return to prison" (SIPU I:9). Nowhere are we told if the referrals were counted as successes, failures, or were dropped from the study. Furthermore, these referrals were primarily for chronic alcoholics and narcotic addicts (SIPU I:9) - a group of parolees who were supposedly screened out of the project (see SIPU I:3).

The results of Phase II were measured by both the Total Arrest Rate and the Parole Violation Rate, since the researchers were unable to decide which was the better measure (SIPU II:22). The difference between Total Arrests and Parole Violations was listed as the Minor Arrest Rate (sentences of less than ninety days). The problem with using both Total Arrest Rate and Parole Violation Rate as criteria can be found when we examine the researcher's definition of "Parole Violation":

Parole Violations are defined as those acts resulting in conviction and sentence of 90 days or more, and, in addition, include all acts following which the parolee is returned to state prison by the Adult Authority. (SIPU II:12)

Since Parole Violations were not broken down to distinguish between "sentences of ninety days or more" and "violations actioned by the Authority", we do not know to what extent Parole Violations, and thus Total Arrests are affected by the parole

agents themselves. Nor do we know the ratio of "ninety days" returns to "Parole Authority" returns between experimental and control groups.

In Phase III the researchers used different criteria (no arrest, minor arrest, major arrest, return to prison) (SIPU III:5). Again, however, the criteria used in this phase present the same problem as did the criteria used in Phase II. The criteria are defined as follows:

Minor arrests are arrests followed by release, acquittal, or sentence of less than 90 days; also included are parole violations not followed by a return to prison. Major arrests are those followed by sentences of 90 days or more and include also cases in which a felony is committed but not prosecuted, cases awaiting trial on a felony charge, and felony probation. Returns include subjects returned on technical violations as well as those returned with new convictions.
(SIPU II;5)

In Phase IV the researchers appear to have partially corrected for the above by clearly distinguishing "Technical Violations" from other measures used (see SIPU IV:12). There is, however, another problem relating to measurement in this phase. The researchers report the loss of 10% of the subjects in one geographic area and 27.8% in another through transfers. We are not told whether or not these "losses" were dropped from the study. This "selection-mortality" problem, of course, represents a serious threat to validity since the experimentals (and possibly controls) are composed of different kinds of subjects

at the time of post-testing (see Cook and Campbell, 1976:227)¹³

THE SIPU STUDY AND CRITERION #3 (ADEQUATE SELECTION TECHNIQUES)

The researcher's plans for each phase of the study was to randomly assign parolees to experimental and control groups from a pool of parolable inmates. While randomization in research is sometimes difficult¹⁴ the SIPU researchers' attempts to randomize suffered from several significant problems. Some of the more obvious deflections from randomization are outlined below.

In the first place, the researchers decided that parolees without employment could be assigned to small caseloads but that those assigned to the regular caseloads had to have employment. Standard practice, prior to the implementation of Phase I,

¹³Cook and Campbell list "mortality" as a threat to internal validity that can be appropriately safeguarded against through randomization. This listing, however, is not correct because although the control and experimental subjects may be similar upon entering the experiment their "reasons" for leaving the experiment might be quite different given the two groups are exposed to different intervention strategies. Not knowing the "reasons" we can not know to what extent they were affected by the impact of the intervention strategy, and thus we can not know the actual effect of the strategy.

¹⁴See Ricken and Boruch (1974:50-67); Kerlinger (1973:323); Cook, Cook and Mark (1977:111-115). While these writings note the difficulties of randomization in research, none express doubt about the feasibility. Boruch, McSweeney, Soderstrom (1978: 655-695) have provided a bibliography of field experiments to illustrate the feasibility of "Randomized Field Experiments". Unfortunately the SIPU Study: Phase III is one of the examples listed by the authors.

required all parolees to have employment upon release. However, when the project began it "was agreed that the increased time each SIPU officer would have available to work with each parolee would make it possible to permit release without employment". (SIPU I;6).

This deflection from randomization carried over into Phase II where "parolees in the experimental group could be released on parole without prior assurance of employment. This is in contrast with the control group where prior assurance of employment was mandatory" (SIPU II:19). An examination of the table on page 21 of the research document shows that while 99% of those on the large caseloads held jobs, only 49% of those on the small caseloads held jobs. (see Appendix C for SIPU Table). There is no indication offered in either the report on Phase III or Phase IV that the issue of employment did not carry over into these later two phases.

On the matter of employment the researchers failed to view the impact of employability as a potentially relevant variable in determining differences in parole success rates regardless of the type of supervision provided. One would assume employable parolees differed from those who were not able to secure employment and that their situations while on parole would be, different enough in important ways, to affect their success rates. One might further assume parolee supervisors spent a certain, if not considerable, amount of time on the intensive

caseloads dealing with matters relating to parolee employment, permitting unaccounted for differences in the type of supervision provided between the intensive and regular caseloads.

A re-assessment of the data presented on page 21 of the research document shows the above assumptions to be most reasonable. The analysis shows that only 14.5% of parolees released to small caseloads were assisted in obtaining employment by a parole officer. On the other hand, analysis of the data shows 55.09% of parolees released to large caseloads were assisted by the parole officer. These figures suggest that parolees assigned to the large case loads may have been more employable and/or cooperative.

We can also see, by re-analyzing data from the Phase II report, that 34.16% of those released to small caseloads were assisted by family or friends in finding employment and only 31.48% were assisted by family or friends without assistance from elsewhere. On the other hand, 43.75% of those released to the large caseloads were assisted by family or friends without assistance from elsewhere. The fact that a greater percentage of the regular caseloads (nearly 10%) received assistance from family and friends might also suggest that a disproportionate number in this group enjoyed greater community support than did some in the small caseloads.

In sum "assurance of employment" as a condition of assignment to the larger caseloads clearly indicates that randomization did not occur and as a result there were important differences between controls and experimentals on employment. Additionally, it is possible that the supervisors on the small caseloads spent a greater amount of time dealing with matters of employment, permitting unaccounted for differences in the type of supervision provided.

Other deflections from the random assignment of parolees to experimental and control groups occurred in each of the four phases. With respect to Phase I, parolees designated as high risk parolees were released to the experimental group only (SIPU I:4). While we are told by the researchers that the evaluation of this group was necessarily excluded and the results tabulated separately, the evaluation of these subjects separately does not necessarily correct for the possible consequences of this deflection from randomization. The "presence" of these subjects in the experimental groups may have had an effect on the other experimental subjects or on the treatment received by the other

subjects.¹⁵

With respect to Phases II and III the researchers themselves inform us of deflections from randomization. They state:

However, at least two factors prevented a purely random assignment of cases without disrupting the necessary functioning of the Placement Program of the Division. First, there were months when for particular areas it was necessary to designate all available cases as experimental. In addition, the late receipt of cases from institutions made it difficult, at times, to insure adequate representation from that prison population. (SIPU II:10; SIPU III:23 - Appendix A)

Additionally, base expectancy scores¹⁶ of Phase III subjects show that controls and experimentals did indeed differ, the scores being significantly higher for the experimentals than for the controls (see Appendix D for SIPU Table). The researchers note, in referring to this difference that:

Although the case assignment method was intended to preclude any such biasing, differences in the time of

¹⁵ These effects, respectively, can be seen as a "history" threat to internal validity and a "random irrelevancy in the experimental setting", and thus a threat to statistical conclusion validity (see Cook and Campbell, 1976:224-233).

¹⁶ A "base expectancy score" or prediction equation is a score assigned to a parolee (or offender) and it is calculated from information on the parolee's background (i.e. record, age) which has been predetermined as significant with respect to recidivism. The score is intended to serve as a predictor of recidivism. The author assumes as do the SIPU researchers that similar scores imply that parolees are similar. It is recognized though that such an assumption is not a good one since it is possible that even if there are similar scores, differences could be masked in the equating process and it is possible that the offenders differ significantly with respect to important, but unknown variables.

reporting parole referrals from the different institutions coupled with the need to keep SIPU caseloads filled and to make assignments before all releases for a given month had been reported made complete randomness difficult to achieve. (SIPU III:24)

The additional deflection from randomization in Phase IV is an obvious one. As well as an attempt to assess the efficacy of reduced caseloads in this phase, the design also involved variation of maturity levels of parolees and the parole agent approach to supervision. The research design called for random assignment of equal numbers of "high and low maturity" parolees to each agent carrying a 15-man, 30-man, or 70-man caseload. A paucity of high-maturity subjects, however, resulted in the assignment of high-maturity subjects to 15-man and 30-man caseloads, and "virtually no high maturity subjects in the large (70-man) caseloads" (SIPU IV:9).

Finally, all four phases of the SIPU study suffer from a deflection from randomization which raises doubt about the external validity of the findings in each phase. Unlike the problems described above, this deflection does not constitute a violation of the present criteria. It is, however, a serious issue and thus merits notation.

The issue is the deflection from randomization with respect to the pool of parolees from which the SIPU subjects were to be randomly assigned. The selection of inmates forming the pool of randomly assigned parolees excluded several kinds of inmates

deemed unsuitable for supervision in intensive caseloads. (see SIPU I:3; SIPU II:5; SIPU III:23; SIPU IV:8). In other words inmates were not considered parolable unless they were suitable for supervision in intensive caseloads. So, contrary to the claims made in each of the four research reports, the SIPU study was not designed to test for the impact of intensive supervision on parolees per se, but, rather, the study was an attempt to assess the impact of intensive supervision on parolees deemed suitable for such supervision. This means that the parolees studied were not a sample of a universe from which one would normally want to generalize (see Cook and Campbell, 1976:235)¹⁷.

THE SIPU STUDY AND CRITERION #4 (ADEQUATE CONTROL OVER INDEPENDENT VARIABLE)

The failure of the researchers to fulfill criteria one and two precludes the conclusion that treatment was received as defined and that only the experimentals received such treatment. It was noted earlier that, at best, we can really only say we don't know who received what. Ignoring problems associated with criteria one and two we are still required to say the researchers failed to meet criterion four, since virtually half of the subjects received about the same number of contacts.

¹⁷Also see Suchman (1967:103); Williams and Evans (1969: 118-132); Kerlinger (1973:324-325); Cook, Cook, and Mark (1977: 108-109); and Rutman (1977:35-36).

Regarding the failure of the study to meet criterion #4, the researchers have tried to make the point that the study did meet it. They believed differences in the intensity of supervision received between experimentals and controls were made evident on the basis of a time study they conducted which revealed that SIPU agents and control agents spent 37% and 40% of their respective time making contacts (SIPU II:9-10). The researchers reasoned from this information that the SIPU agents must have spent three times as much time per case in contacts as did the control agents (SIPU II:10). Here again, the researchers are using a frequency/time concept to indicate an essentially qualitative concept. Assuming we are willing to overlook this problem, the time study does not rule out the very likely possibility that the length of contacts varied within wide limits for both groups and further, that a large number of parolees received the same number of minutes in contacts in the small as in the large caseloads. In any case, the length of time for contacts was not measured. We might also bear in mind that supervision was to be measured in terms of "frequency", not "length of time in contacts". Thus a conclusion that experimentals and controls received different treatment is unwarranted. The researchers themselves support this conclusion in reporting on the findings of Phase II when they state:

...the lack of clear-cut differences in frequency of contacts between the experiential and control groups may have helped to obscure whatever real differences might

~

have existed between these two groups.
(SIPU II:1)

THE SIPU STUDY AND CRITERION #5 (ADEQUATE RESEARCH REPORT)

The research reports on each phase of the study failed to provide sufficient information necessary for the discovery of possible error in both the data and research design. The assessment of the study's methodological adequacy was, in many instances, hampered by the unavailability of data and other important information, and thus verification was impossible. Moreover, the lack of specific information on selection procedures, and the definition and measurement of intensive supervision make replication of each phase of the study impossible..

In fairness to the researchers, it should be realized that the reports on each phase of the study did contain a significant amount of detailed information. It appears the researchers' intent was to provide a comprehensive report on each phase. Perhaps more importantly, it appears the researchers made an attempt to present all information in an honest and forthright manner.. Throughout each report the researchers draw attention to problems which they believed could have influenced the program and its outcome. The researchers themselves cite several visible problems not already discussed including the failure to randomly assign parolee supervisors to large and small caseloads (SIPU

I:5; SIPU II:6; SIPU III:2; SIPU IV:4); personnel turnover (SIPU I:9; SIPU II:6; SIPU IV:8); and masking effects (SIPU IV:1).¹⁸

In addition to the above disclaimers, the researchers both summarize the evaluations of Phase I by a consultant, Dr. N. Rudy, Professor of Statistics and publish his entire evaluation in an appendix. Dr. Rudy summarized his evaluation by stating that the results of the SIPU study were "largely inconclusive" (see SIPU I:29-34).

SUMMARY AND CONCLUSIONS

A detailed and careful examination of the entire four phases of the SIPU project as recorded in the original research documents reveals that none of the phases satisfied any one of the criteria used here. The examination has shown at least that parolees were not randomly assigned; the independent variable (intensive supervision) was not adequately defined or measured; the dependent variable was inadequately measured and reported; there was no evidence that experimentals received treatment as defined, and that the controls did not receive treatment; and the research reports on each phase of the research failed to provide information essential to the study's replication and verification. These significant inadequacies, clearly suggest

¹⁸ "Masking effects" means the intervention strategy (treatment) may have had opposite effects upon different kinds of subject within the experimental groups. Thus an effect may be covered up (masked) or cancelled out in the summing up of the data process.

that each phase of the study is so poor there are no conclusions to be drawn from them singly or in combination in terms of substantive findings or their social policy implications for criminal justice.

As noted in the preceding chapter more than fifty different social scientists have cited the SIPU research. The large majority of these citations refer to the original research documents. Of all citations to the SIPU study discovered more than fifty cite the original research documents. (see Tables II and III, chapter 3). Despite this great number of references to the original research documents, there remain only three instances where attention has been given to specific methodological problems (e.g. Waller, 1972: 127; Logan 1972:378-387; Lipton, Martinson, and Wilks, 1975:116-163).

Of the three instances where attention has been given to methodological problems, only one (Logan, 1972:378-387) found the study's methodology severely flawed.¹⁹ However, even Logan's analysis was not entirely correct as he reported the researchers randomized when they did not (see Logan, 1972:386-387). Waller (1972:116-205) referring to Phase III and IV, also erred in that

¹⁹ Logan (1972:386-387), in an assessment of the first three phases, found that the study failed to provide an adequate definition of intensive supervision; failed to provide evidence that only the treatment group received treatment; and was incapable of routinization.

his methodological criticism is only of Phase III.²⁰

The failure of Waller (1972) and Logan (1972) to deal more completely with methodological issues can be understood when we consider both observers failed to consult the original document in its entirety.²¹ These writers, therefore, did not have access to all the facts. The failure of Lipton et al. (1975) to regard the SIPU research as being methodologically unsound when, in fact, it met virtually all of their criteria for exclusion is more difficult to understand. Indeed, as mentioned in the preceding chapter, the research was classified as four of the best of the many studies they had reviewed (see Lipton et al, 1975:153,156,160,161). This is particularly confusing as Lipton et al note that the introduction of small caseloads was not put into effect and therefore did not differ from the regular caseload supervision (Lipton et al, 1975:119). In addition, they point out in a footnote that there was "an absence of either a theory of supervision or any measure of 'adequacy' of supervision" (Lipton et al, 1975:119). The authors also point to some of the limitations of the outcome measure (Lipton et al, 1975;136-137). These problems alone clearly satisfy their

²⁰ Waller (1972:127) noted that the researchers failed to check the nature of the parole officers' decision to revoke parole.

²¹ Logan (1972) cites Glaser (1964:457-8) as reference to Phase III and thus was relying on secondhand information about this phase, while Waller (1972) relied on Phases III and IV as a reference to information about the first two phases.

criteria four, five, six, and nine for exclusion²².

The more than fifty other occasions where the SIPU research has been cited, and taken at face value without attention to methodological issues appear to constitute a clear cut instance of the failure of criminologists to adhere to the prescription of "organized scepticism". Even conceding that some criminologists did employ "scepticism", it seems clear, particularly in light of the cases where attention had been given to methodological concerns, the process itself is not a reliable tool for ensuring the scientific adequacy of research.

While it has been found that conformity to methodological norms is variable among social scientists, or at least among evaluative researchers, (Bernstein, 1978:24-47)²³ this finding cannot be used to explain the findings of the present analysis. If it could serve as even a plausible explanation, we would expect to find at least some variance in the citations to the

²² One would only be guessing in an attempt to explain the confusing conclusions of these authors. One of the authors, though, has been criticized elsewhere regarding the presentation of his work. Thus Adams (1977:325) states "careful examination of Martinson's methods and conclusions suggests that he has given us a mixture of science, rhetoric, and legerdemain, and it is difficult to say what predominates".

²³ Bernstein's study suggested that this variance can be partially explained when comparing social scientists who are guided by the norms of an "entrepreneurial model" (more likely in evaluative research) to those scientists who are guided by the norms of an "academic model". The study indicated that one of the consequences of the "entrepreneurial model" is the lower conformity to the prescribed norms of methodology in social science.

SIPU study with respect to its methodological adequacy. That is to say we should have found among the more than fifty citations to the SIPU study at least some evidence of awareness and mention of methodological flaws to lesser and greater degrees. Again, such awareness and mention is all but totally absent.

III. MISREPORTING

INTRODUCTION

The preceding chapter showed the SIPU study to be so methodologically flawed its findings must be rendered nugatory. Only once (i.e. Logan, 1972:380) in more than fifty citations has a similar observation been offered in the literature. The absence of such an observation being reported appeared to be a result the failure of criminologists to exercise the norm of "organized scepticism", the unreliability of the process of organized scepticism itself, or both. There are, however, peculiarities among the citations to the SIPU studies not mentioned in the previous chapter, which suggest the problem of such massive error being overlooked is not only the result of "failed" scepticism, but additionally the result of other factors.

The possibility the error might be the result of other factors came to mind when, in considering more closely those writings in which "general"¹ methodological concerns were raised, it seemed peculiar that the reporter would criticize

¹ "General" meaning that mention was not given to a specific methodological problem or the degree to which inappropriate methods affected the study.

certain phases of the study and not others. At the same time, several citations provided incorrect information about the SIPU findings. Moreover, it was noticed this misreporting occurred regardless of whether the author of the citation referred to the original research document or cited a secondary source.

The above cursory observations led to a conjecture that additional insights to why the SIPU study did not receive critical methodological reappraisal might be provided by a more thorough analysis of the citations themselves. At the very least such an analysis might show the extent of "misreporting". Clearly, since misreporting of SIPU findings appeared to be the rule rather than the exception a more illuminating analysis would be provided by examining all the available writing which reported on the entire four phases of the SIPU research rather than just studying those few writings which gave attention to methodological concerns. This chapter presents and discusses the analysis and findings.

The presentation begins with a general statement of the findings and then explains how the analysis sample was selected. This explanation includes a table providing information on excluded citations. A second table contains all of the information from which the analysis is made. A statement is then provided on the SIPU findings as reported in the original research document. Finally, the information, is analyzed and some conjectures are offered.

FINDINGS IN BRIEF

A review of a selected number of reports clearly shows that the SIPU study has been widely misrepresented. Among seventeen reports, only one reported on the findings of the study correctly. Among the sixteen incorrect reports there is a total of eight different versions concerning what was actually reported by the SIPU researchers. The discrepancies between versions were about the alleged findings of Phase III and to a lesser degree, Phase IV.

There were also several cases of incorrect information in descriptions of the study, poor descriptions of the study in the majority of reports, incomplete (did not report on all phases) descriptions in many reports published ten or more years after the publication of the original SIPU report, cases of inappropriate terminology describing "statistically significant" findings, and cases of discrediting statements applied only against selected phases of the study.

Apparently, many reports relied on secondary or "para-secondary"² sources for information about the SIPU study. Analysis shows, however, that while secondary and para-secondary sources are perhaps a contributing factor in some misreporting,

² "Para-secondary" means a reporter used the report of one phase of the study as a source of information to report on other phases.

secondary sources ~~per se~~ did not appear in most cases to be a plausible explanation for misreporting. Also, most misreporting could not reasonably be attributed to the presence of multiple conclusions or the manner in which the findings were presented in the original research report since the majority of citations reported findings which were either entirely opposite to the findings the SIPU researchers reported or which were not even mentioned in the original report. Finally, there is no apparent pattern to the use of inappropriate terminology and selectively applied discrediting statements, and such uses did not support the notion of treatment destruction techniques as recently put forth by Gottfredson (1979:39-54).

SELECTING A SAMPLE

Selecting a sample for the present analysis began with a search through the University of British Columbia and Simon Fraser University libraries for all writings which made reference to the SIPU study. This search located fifty-five citations³.

There likely are other citations buried in reports not supplying references, included in unpublished manuscripts, or

³ It was surprising to find that most of the citations found in the libraries were not listed in the "citation index" (i.e. Social Sciences Citation Index) The search was made by looking through the "stacks" for books and journal articles which seemed likely (by virtue of their titles) to cite research on caseload size.

contained in writings simply not available in the libraries at U.B.C. and S.F.U. For example, the eighth and ninth editions of Sutherland and Cressey's text Criminology could not be located, but we can be reasonably sure the SIPU study would be mentioned in these editions since it is mentioned in the seventh and tenth editions. It is assumed, though, that these unavailable or undetected citations are small in number and their exclusion from the present analysis is not significant.

Of the fifty-five reports located, seventeen were chosen for further study. Reports were chosen only if they cited all four phases of the SIPU research or if they gave a clear impression they were making reference to each of the four phases. Thus, reports such as those of Reckless (1961:546-547), Gibbons (1968:529), Miller (1979:256-262), and Dressler (1969:268-269) were excluded because they make reference to only one or two phases of the study. Similarly, LeDain Commission's report (1973:1043,1055-1056) was excluded since it was impossible to determine whether the commission was referring to all four phases or simply Phase IV - the phase cited as reference.

It also had to be obvious that reporter's were only reporting the findings of the SIPU researchers and not relaying their own interpretation. Since the present analysis is concerned with how accurately the SIPU has been reported and not how the study has been interpreted, the slightest doubt about

any part of a discussion on the SIPU findings resulted in a report's exclusion. Thus, for example, Keve (1967:42), Adams and Vettors (1971:392-393), and Lipton, Martinson, and Wilks (1975:116-163) could not be included. Lipton et al (1975) was the only report excluded on interpretation grounds alone.

Also excluded were citations such as those of Robinson (1969:50-51,126), Outerbridge (1970:192,199), and Adams (1974:1028-1029 which offer explanations of the SIPU study in reports already included among the seventeen selected for further study. In other words, these reports are duplicates, and thus their inclusion would not add to the analysis. Finally, it was necessary to exclude any report which failed to supply a reference. Thus, the reports of Conrad (1965:295) and Carney (1977:106) were excluded. A complete list of excluded reports and the reasons for their exclusion is provided in Table II. Table III lists inclusions.

TABLE II

CITATIONS EXCLUDED FROM THE ANALYSIS

#	NAME OF AUTHOR(S) CITING SIPU RESEARCH	INFORMATION REPORTED				REFERENCE USED BY AUTHOR(S)	REASON FOR EXCLUSION
		Phase I	Phase II	Phase III	Phase IV		
1	Conrad, 1965			General Statement Positive Findings	No reference	No reference	
2	Carney, 1977			General Statement Negative Findings	No reference	No reference	
3	Gibbons (1st.ed.) 1968	N	N		Reimer & Warren, SIPU I	Incomplete	
4	Gibbons (2nd.ed.) 1973	N	N		Reimer & Warren, SIPU I	Incomplete	
5	Reckless (3rd.ed.) 1961 *	N	N		SIPU II	Incomplete	
6	Reckless (4th.ed.) 1967 *	N	N		SIPU II	Incomplete	
7	Reckless (5th.ed.) 1963 *	N	N		SIPU II	Incomplete	
8	O'Leary, 1969 *	N	N		SIPU I & II	Incomplete	
9	Tappan, 1960	N	N		Reimer & Warren,	Incomplete	
10	Cavan (3rd.ed.) 1962	N	N		Reimer & Warren,	Incomplete	
11	Miller, 1979	P	P		SIPU I & II	Incomplete	
12	Sutherland & Cressey (7th.ed.) 1960	N	N	0	Reimer & Warren, SIPU II & IV	Incomplete	
13	Sutherland & Cressey (10th.ed.) 1978	N	N	0	Reimer & Warren, SIPU II & III	Incomplete	

TABLE II (continued)

#	NAME OF AUTHOR(S) CITING SIPU RESEARCH	INFORMATION REPORTED				REFERENCE USED BY AUTHOR(S)	REASON FOR EXCLUSION
		Phase I	Phase II	Phase III	Phase IV		
14	Dressler (2nd.ed.) 1969			P&DE		SIPU III	Incomplete
15	Logan, 1972	N(M)	N(M)	N(M)		Reimer & Warren, Glaser (1964)	Incomplete
16	The Ledain Commission, 1973				N	SIPU IV	Incomplete
17	Robinson, 1969	N	N	P&DE	N/N	Adams (1967)	Incomplete
18	Outerbridge, 1970	N	N	P&DE	N	Original	Duplicate
19	Adams, 1974	N	N	P&DE	N	Original	Duplicate
20	Glaser, 1964	N	N	DE	O	Reimer & Warren, SIPU III	Duplicate
21	Glaser (Abridged ed.) 1969 (a)	N	N	DE		SIPU III	Incomplete
22	Glaser, 1969 (b)	N	N	DE	??	Reimer & Warren, SIPU III, Reported	Incomplete
23	Lipton, Martinson, Wilks, 1975	?(M)	?(M)	?(M)	?(M)	Original	Interpretation & Incomplete
24	Adams & Veeters, 1971			?		SIPU III	Interpretation
25	Keve (3rd printing 1973) 1967*	?	?	?		SIPU III	Interpretation & Incomplete
26	McCleary, 1975				?	SIPU III & IV	Interpretation & Incomplete
27	Gibbons, 1969	?				SIPU I	Interpretation & Incomplete

TABLE II (continued)

#	NAME OF AUTHOR(S) CITING SIPU RESEARCH	INFORMATION REPORTED				REFERENCE USED BY AUTHOR(S)	REASON FOR EXCLUSION
		Phase I	Phase II	Phase III	Phase IV		
28	Bean, 1976			N		SIPU III	Incomplete
29	Clegg (3rd printing 1974) 1964						
30	Brill, 1978			P		SIPU III	Incomplete
31	Adams, 1976				DE	SIPU IV	Incomplete
32	(Internal Memo) National Parole Service, 1978			O	O	Original	Duplicate
33	Maltz & McCleary, 1977					Lipton et al. (1975)	Not specific about each phase
34	Waller, 1974				O	SIPU IV	Incomplete
35	Gottfredson, 1975				N	SIPU III & IV	Incomplete
36	Boruch, McSweeney, Soderstrom, 1978				O	SIPU III & IV	Incomplete
37	Barnes & Teeters (3rd.ed.) 1959				O	SIPU IV	Incomplete
38	Manual of Corr. Standards, 1966					Reimer & Warren, SIPU IV	Incomplete

TABLE III

INCLUDED CITATIONS

#	NAME OF AUTHOR(S) CITING SIPU RESEARCH	INFORMATION REPORTED				REFERENCE USED BY AUTHOR(S)	CRITICISM NOTED OR MADE BY AUTHOR(S)
		Phase I	Phase II	Phase III	Phase IV		
1	Outerbridge, 1968	N	N	N	N/N	SIPU IV	
2	Kassenbaum, Ward, Wilner, 1975	N	N	N	/N	Original	
3	Robinson & Smith, 1971	N	N	DE	N/N	Adams (1967)	Phase III & IV
4	Klapmuts, 1973	N	N	DE	N/N	Original	
5	O'Leary, 1974	N	N	DE	N/N	SIPU III & IV	Phase IV
6	Adams, 1967	N	N	P&DE	N/N	Original	Phase IV
7	Duffee & Fitch, 1976	N	N	P&DE	N/N	Original	
8	Carter & Wilkins, 1968	N	N	P&DE	N/N	Adams (1967)	
9	The Citizens Inquiry, 1975	N	N	P&DE	N/N	Robinson & Smith (1971)	
10	Burkhart, 1976	N	N	P&DE	IN/IN	Original	
11	Lerner, 1977	N	N	DET	/N	Original	
12	Martinson, 1974	N	N	P	N/N	Original	Phase III
13	Public Health, 1971	N	N	DE	/DE	Original	
14	Harlow, Weber, Wilkins, 1971	N	N	DE	/DE	Original	
15	Soloman, 1976	N	N	DE	/DE	Original	
16	Waller, 1972	N	N	P(M)	/DE	SIPU III & IV	Phase III & IV
17	Greenberg, 1977	N	N	P&DE	/P	Robinson & Smith (1971) SIPU IV, Adams (1967), Lipton et al (1975)	Phase III

LEGEND FOR TABLES II, III, IV and V

NOTE: When space is blank it means that the phase was not reported on. With respect to the column for Phase IV, the top symbol refers to findings on parolee/parole officer matching while the bottom symbol refers to findings on caseload reduction.

N - Author reports negative findings found

P - Author reports positive findings found

DE - Author reports differential effects found

DET - Author reports deterrent effects found

IN - Author reports that findings were found to be inconclusive

P&DE - Author reports positive findings which include differential effects

* - Author provides incorrect information on design of research

O - Author does not discuss findings regarding caseload size, although the research is cited for other reasons (i.e. information on advanced release, prediction techniques)

(M) - Author provides specific methodological criticism

? - Difficult to determine whether or not author is offering own interpretation

?? - Difficult to determine what author is stating about findings

Incomplete - Fewer than all four phases were cited

Duplicate - Author has another article which is included in Table II

FINDINGS AS REPORTED IN THE ORIGINAL REPORT

Given the nature of this analysis it is obviously essential to know precisely what was stated in the original report.

Therefore, extensive quotations from the original SIPU research document are included. A reading of the four SIPU reports reveals the following findings being reported:

Phase I (1956). From page ten of the research document two statements show that negative findings were reported in this phase:

One of the suggestive statistics that came out of the 1955 SIPU report was that SIPU had a differential effect which was favourable to robbers, forgers, and burglars, especially recidivist robbers (see Appendix 111). The analysis of later statistical data reveals that these differences no longer apply. (SIPU I:10)

While the differences in violation rates between the 15 man and 90 man caseloads were small and insignificant the violation rates have remained favourable to the intensively supervised 15 man caseloads. (SIPU I:10)

Phase II (1958). Under the heading "Summary of Findings" on page one of the research report a clear statement of negative findings can be found:

~~There was no significant difference between the thirty man and ninety man caseloads in recidivism criteria (19.45) Parole Violation Rate to (20.4%) favoring the thirty man caseloads. (SIPU II:1)~~

Phase III (1962). The SIPU researchers offer a statement of positive findings generally and a finding of a differential

effect SIPU researchers on page one of the research report:

The findings can be summarized as follows:

1. Parolees released to reduced caseloads performed significantly better than those released to regular caseloads, at both twelve and twenty-four months after release.
2. The difference between SIPU and Control caseloads was not the same for all parolees.
 - a. The SIPU-Control difference was larger for medium-risk parolees than for the best or poorest risks.
 - b. The SIPU-Control difference was larger for parolees released to the northern region than for the southern region.
 - c. The SIPU-Control difference was larger for parolees released in the middle of the time period covered by the study than for parolees released earlier or later.
(SIPU III;1)

Phase IV (1965). In this phase two distinctly different hypothesis were tested. Concerning the hypothesis that an appropriate matching of parolee and parole agent types can reduce parole recidivism the research document states:

The Characteristics of parolee - parole agent relationships studied so far have been shown to have little to do with parole outcome. In this project unforeseen research and administrative factors lead the staff to believe that an adequate test of the matrix relating to parolee - agent interaction has not occurred... The hypothesis about the interaction matrix merits further testing under more exacting experimental conditions and an examination of the processes involved.
(SIPU IV;ii)

The predicted parolee-parole agent interaction did not occur. Neither high nor low - maturity parolees performed differently when supervised by external, as compared to internal agents. (SIPU IV;ii)

On the hypothesis concerned with caseload size and parole outcome relationship the report states:

Subjects in small (15 man) and in medium (30 man) caseloads performed significantly better than subjects in large (70 man) caseloads in terms of proportion of no arrests and of minor arrests. Differences in proportion of major arrests, TFT returns, and WNT returns were not significant, nor were there any differences in proportion of major violations or total returns to prison... There was no indication that small (15 man) caseloads were more effective than medium (30 man) caseloads with subjects of any risk level. (SIPU IV;iii)

What may be concluded from the findings reported on caseload size? SIPU IV confirms the SIPU III finding of a medium (30 man) vs. large (70 man) caseload difference. Medium (30 man) caseloads in the SIPU IV study performed significantly better than large (70 man) caseloads. Their superiority was most evident in keeping men arrest free for the first year of parole and in preventing arrests of a minor nature. (SIPU IV:25)

If we consider the SIPU study as as one report, we see the SIPU researchers reported negative findings in Phase I and II, positive findings and a finding of a differential effect in Phase III, positive findings in Phase IV with respect to the efficacy of reduced caseloads; and inconclusive findings with respect to the testing of parolee - parole agent interactions in Phase IV.(Listed in Table IV).

"Differential effect" means the treatment had a greater effect on some offenders than on others. "Positive findings and a differential effect" means the treatment produced a significant difference overall, but the effect was different depending on the type of offender. It does not mean the differences between caseload sizes were only significant for certain groups.

DIFFERENT VERSIONS

A comparison of the seventeen citations on Table IV to the report of the SIPU researchers shows that only once have the findings of the SIPU study been reported correctly (i.e. Greenberg, 1977:135-136). Indeed, among incorrect reports there are eight different versions concerning the SIPU research. Table III shows, though, that each citation reports the first two phases correctly and the discrepancies are only about the last two phases. There are five versions concerning Phase III and four versions concerning Phase IV. On the basis of what we know about differences between the original reports on the first two phases and the last two, we might conjecture these discrepancies are attributable to two closely related factors.

TABLE IV

REPORTS ON SIPU FINDINGS

#	VERSION OFFERED				AUTHOR(S) REPORTING
	Phase I	Phase II	Phase III	Phase IV	
1	N	N	N	N	Outerbridge, 1968; Kassebaum, Ward, Wilner, 1975;
2	N	N	DE	N/N	Robinson & Smith, 1971; Klapmuts, 1973; O'Leary, 1974;
3	N	N	P&DE	N/N	The Citizens Inquiry, 1968; Carter & Wilkins, 1968; Duffee & Fitch, 1976; Adams, 1967;
4	N	N	P&DE	IN/IN	Burkhart, 1976;
5	N	N	DET	/N	Lerner, 1977;
6	N	N	P	N/N	Martinson, 1974;
7	N	N	DE	/DE	Harlow, Weber, Wilkins, 1971; Public Health, 1971; Solomon, 1976;
8	N	N	P	/DE	Waller, 1972;
9	N	N	P&DE	/P	Greenberg, 1977; SIPU Researchers;
TOTAL	1	1	5	4	

MULTIPLE CONCLUSIONS AND A POOR PRESENTATION

First, the reports on the last two phases seem to provide multiple conclusions, while the reports on the first two phases provide only one conclusion each. Actually, the last two phases each only provide one conclusion; unless read carefully, however, only half of the conclusion would be reported. Second, the discrepancies may be due to findings of the last two phases not being presented as clearly as they could have been. For example, the summary of findings in the Phase III report tells us the differences between SIPU and Control caseloads was not the same for all parolees, but it does not tell us if these differences were insignificant with respect to any one group. Therefore, we are left to assume, if we do not read elsewhere in the report, that the findings were significant with respect to all parolees. Similarly, the summary of findings in the Phase IV document reports findings on the testing of parolee - parole agent interaction as negative, when in fact clearly the researchers considered these findings as nugatory, or at least inconclusive.

One problem, however, in attributing misreports in this case to either the presence of multiple conclusions or poorly presented findings, or even a combination of the two, is that we

are left to consider why an author would cite a finding not even offered in the original report. All sixteen of the misreports cited findings concerning caseloads sizes in Phase IV which were not mentioned in the original report. This was also the case for three citations on the alleged findings of Phase III.

Moving from Table III momentarily to Table II we notice a citation which makes a claim about Phases I and II not offered in the original report:

In fact, research studies have indicated that the less contact the parolee has with the parole officer, the better the likelihood of success in the community. (emphasis added) (Miller, 1979:256-257)

The original reports of Phases I and II are cited as examples of "the more notable studies proving this point" (see Miller, 1979: 262).⁵

SECONDARY SOURCES

While all citations referred to in this analysis are secondary (or tertiary)⁶ sources to any reporter who might rely on them for information about the SIPU study, they, with the exception of Greenberg (1977:135-136), contribute to misreporting. Additionally, Table III shows that seven of the

⁵ This citation, however, is arguably Miller's own interpretation and not a "report" of the research.

⁶ A "tertiary source" means a source which relies on a secondary source to report on the SIPU study.

sixteen incorrect citations relied on secondary or para-secondary sources. While this may create the impression that secondary and para-secondary sources contribute to misreporting a closer inspection shows, however, that a reliance on secondary (including para-secondary) sources per se does not appear to explain misreporting in this case. Indeed, only three of the citations referred to secondary sources. In the four cases in which the citation was based on a para-secondary source, misreporting of findings only occurred when the original source was used. One citation, Outerbridge (1969:379-380,386), using Phase IV as reference, did misreport the findings of Phase III. In a later writing, however, Outerbridge misreported these findings when using all four phases as reference (see Outerbridge, 1970:192,199). It is noteworthy that the only correct report relied on a para-secondary source.

Further, removing secondary sources from Table III does not decrease the number of incorrect versions from eight nor does it change the number of incorrect versions concerning Phases III and IV.⁷

⁷ Also, it is reasonable to assume that if secondary sources were indeed responsible for misreporting we would expect tertiary citations to report the same conclusion as the secondary source. We can see from Table III that this is not always the case - Robinson and Smith (1971:76-77) report a differential effect for Phase III when their reference, Adams (1967:48-49), reports positive findings and a differential effect for this phase. Looking at Table II we can see yet another example; Logan's (1972) use of Glaser (1964) for reference. Logan reported a negative finding with respect to Phase III, while Glaser reports a finding of a differential

INAPPROPRIATE LANGUAGE AND SELECTIVE CRITICISM

Finding reporters using inappropriate language to describe "significant" findings and applying discrediting statements against only certain phases of the study is a cause for concern. While these problems are not necessarily problems of misreporting, the latter problem does seem to suggest that some social scientists may be selectively sceptical in examining original work [see Barnes and Dolby (1970:10) Footnote, 2 Introduction Chapter I above]⁸. Also, certainly both problems as they apply to the SIPU study remind us of the notion of "treatment destruction techniques"⁹ as forwarded by Gottfredson (1979:39-54). For these reasons it seems necessary to examine -----
⁷(cont'd) effect in this phase.

⁸ Indeed one is also reminded of Miller's (1977:481) notion of "ideologized selectivity". Miller states that "Ideology plays a paramount but largely unrecognized role in this process, to the detriment of other ways of determining priorities. Ideologized selectivity exerts a constant influence in determining which problem areas are granted greatest significance, in which projects are supported, what kinds of information are gathered and how research results are analyzed and interpreted". (Miller, 1977:481) (emphasis added) This is only mentioned here as "food for thought", the author is not claiming to know the motives or the ideological stance of any scientist referred to in this thesis.

⁹ The techniques Gottfredson lists include: (1) contaminate the treatment; (2) stress the criterion problem (attack the criteria used as an indicator of illegal behavior); (3) appeal to common sense; (4) demonstrate that rehabilitation is premised on faulty theory; and (5) seek universals (see Gottfredson, 1979:39-54).

them. Gottfredson argues that:

...there are at least five distinguishable methods that may be used to demonstrate the ineffectiveness of any and all treatment modalities in the criminal justice system. Individually (but, most effectively in combination), these methods are capable of destroying any positive results that might appear in the literature. Perhaps more important, each can be used to show that continuing research in the area would be a mistake...many of them are already an integral part of the working vocabulary of many criminologists. Gottfredson, (1979:40)

Clearly, each of the three criticisms levied against the findings of Phase III are examples of what Gottfredson calls the "contaminate the treatment" technique whereby the critic suggests "that the effect reported in the study is due to some treatment other than suggested by the authors of the study"¹⁰. Thus, Martinson (1974) cites positive findings for Phase III, but attributes these findings to a deterrent effect (see Martinson, 1974:46-47). Robinson and Smith (1971:76-77) and Greenberg (1977:135-136) cite positive findings for Phase III, but add there was "evidence" to show the results were not a result of the effectiveness of reduced caseloads, but instead the results of differential decision making that originated from a desire to demonstrate the effectiveness of small caseloads. In the case of Robinson and Smith (1971) and Martinson (1974)

¹⁰ Gottfredson states that for maximum impact "the technique should be invoked only after a review of the study has appeared to show a substantial treatment effect. Let the reader grasp this ray of light in an otherwise darkened room of negative findings. Suddenly snap the windowshade shut by contaminating the treatment (Gottfredson, 1979:40-42)."

negative findings were reported in the other three phases.

Greenberg (1977) reports positive findings in Phase IV as well but presents them in a quite negative way:

Although there were small and significant differences in the proportions of each group experiencing no arrest or minor arrests, there were no significant differences in terms of major arrests, technical revocations of parole, or returns to prison with a new commitment. (emphasis added) (Greenberg 1977:136)

While Gottfredson's techniques do not include the use of semantic legerdemain, including the use of inappropriate terminology, we can suppose such use could provide a similar effect. From the citations listed on Table III we will find three instances where inappropriate terminology was used, negative findings had been reported in all other phases. Thus instead of the word "significant" O'Leary (1974:937) reports "some" differential effects for Phase III, while Robinson and Smith (1971:76) report "slightly better" differential effects for this phase. The Citizen's Inquiry Committee (1975:170) also uses the term "slightly better" in reporting positive findings and a differential effect for Phase III.

The failure of these three citations to use appropriate terminology and the use of discrediting statements in the three previous mentioned citations means, of course, that the findings reported by all six could well be taken by subsequent observers as being negative, or at least not worthy of further attention. This, together with the fact that discrediting statements could

have been levied against the "negative" findings for one reason or another, is consistent with the notion of treatment destruction techniques. On the other hand, to charge that the authors deliberately employed treatment destruction techniques implies a knowledge of each reporter's motives, a knowledge we don't have.¹¹ Further there are still three instances where discrediting statements were levied against negative findings. Thus Adams (1967:49), O'Leary (1974:937), and Robinson and Smith (1971:77) note in one sentence that Phase IV had been criticized for "a lack of precision". Indeed, in two of these instances (i.e. Robinson and Smith, 1971; O'Leary, 1974) the discrediting statements are offered by the same reporters we would be accusing of employing treatment destruction techniques.

¹¹ Not surprisingly, Gottfredson was recently criticized for claiming (at least implying) to know how social scientists evaluate effectiveness claims. Thus Lerman (1980:126-127) correctly states "If it could be demonstrated that a panel of reviewers from the National Academy of Sciences, as well as other social scientists and criminologists, were not directly reading studies and using scientific criteria to evaluate effectiveness claims, but instead were only relying on the secondary analyses of the inventors of destruction techniques, then this would constitute important news. However, Gottfredson does not provide any evidence that any of the authors cited in his opening paragraphs reached their no-impact conclusions by relying solely on the authors of the alleged 'destruction techniques'. Unless he is willing to make, and sustain, this kind of charge, there is really no basis for suggesting that current conventional wisdom has not been employing traditional standards of scientific proof." Gottfredson, in a rejoinder to Lerman, clearly avoids this request for "evidence" (see Gottfredson, 1980:130-122).

INADEQUATE DESCRIPTIONS, INACCURATE BASIC INFORMATION,
INCOMPLETE REPORTS

Analyzing citations to the SIPU study shows that not only the findings of the study have been poorly reported. The majority of reports, included and excluded, provided inadequate descriptions of the SIPU study. The original report tells us (as noted in Chapter II) that in Phase III there were only two caseload sizes - a thirty-five man caseload and a seventy-two man caseload. Some reports fail to mention that caseload sizes for controls in Phase III had been changed from ninety men to seventy-two men. Others provide entirely inaccurate statements concerning basic information. Outerbridge (1968:380), for example, reports that caseload sizes for Phase III were set at fifteen, thirty, and seventy, while Glaser (1969(b):23) reports that caseload sizes were set at thirty-five and ninety. O'Leary (1969:215) provides incorrect information in his bibliography by referring to a twenty man caseload study. Outerbridge (1968) also erred by noting that parolees on intensive caseloads were given psychiatric consultation (see Outerbridge, 1968:379). There is no mention of such treatment anywhere in the Phase I report, although the Phase II report, in providing an introductory statement on Phase III, does state that in the third phase a psychiatrist and a psychiatric social worker did meet the agents of the Los Angeles Unit on a bi-weekly basis

(see SIPU II:24).

It can readily be seen by looking at Table II that most of the excluded reports were incomplete. Some of these writings [e.g. Glaser, 1969(a), 1969(b); Gibbons, 1968; O'Leary, 1969; Reckless, 1967] were printed in the late sixties and thus their authors might not have had access to the reports of the entire four phases of the SIPU study. Such writings cannot be faulted for failing to provide information on all four phases. There are, however, cases where access should not have been a problem. For example, the fifth edition (1973) of Reckless's text The Crime Problem and the second edition (1973) of Gibbons' text, An Introduction to Criminology were not updated to include statements on Phase III and IV findings. Similarly, Dressler (1969:268) cited only Phase III, while Miller (1979:262) cites only Phase I and II. The LeDain Commission Report (1973:1043,1055-1056) cites only Phase IV, and Sutherland and Cressey (1978:640-651) fail to refer to Phase III. While these incomplete reports are not necessarily related to problems of misreporting, in as much as they fail to provide the entire picture when doing so would be relevant, they are possible sources of distortion. Nonetheless, as shown by Table V below, among incomplete reports there are several accurate references to the SIPU findings.

TABLE V

CORRECT INCOMPLETE CITATIONS

#	NAME OF AUTHOR(S) CITING SIPU RESEARCH	INFORMATION REPORTED				REFERENCE USED BY AUTHOR(S)
		Phase I	Phase II	Phase III	Phase IV	
1	Gibbons (1st.ed.) 1968	N	N			Reimer & Warren, - SIPU I
2	Gibbons (2nd.ed.) 1973	N	N			Reimer & Warren, - SIPU I
3	Reckless (3rd.ed.) 1961*	N	N			SIPU II
4	Reckless (4th.ed.) 1967*	N	N			SIPU II
5	Reckless (5th.ed.) 1963*	N	N			SIPU II
6	O'Leary, 1969 *	N	N			SIPU I & II
7	Tappan, 1960	N	N			Reimer & Warren,
8	Cavan (3rd.ed.) 1962	N	N			Reimer & Warren,
9	Sutherland & Cressey (7th.ed.) 1960	N	N		O	Reimer & Warren, - SIPU II & IV
10	Sutherland & Cressey (10th.ed.) 1978	N	N		O	Reimer & Warren, - SIPU II & III
11	Dressler (2nd.ed.) 1969			P/DE		SIPU III

CONCLUSIONS

The extent to which the SIPU study has been misreported is disquieting. With only one instance of correct reporting of the study when all four phases are reported together, there can be no doubt the problem is something more than one of "failed scepticism".

There is the possibility that one of the cases reviewed here (ie. Kassebaum, Ward and Wilner, 1971:308), although citing all four phases of the study, was really only referring to the findings of Phase I and II. In this case, although each of the four phases was listed as reference, only a general statement was offered. The findings of each phase were not considered separately. Thus, this citation, is arguably, not misreporting - the other citations leave absolutely no room for doubt as to what they are referring to.

Only the reporters themselves can tell us why misreporting occurred in each case. We can only conjecture. Perhaps in some cases the fault lies with the practice of relying on secondary sources instead of the original report. Perhaps it is possible that a reporter cited the original, but actually used a

secondary source. Perhaps in some cases the notion of treatment destruction techniques can explain why a reporter would only cite one phase, criticize one phase, or offer a finding not offered in the original report. The examination provided here, however, has shown, at least, that it is unlikely misreporting resulted from either a reliance on secondary sources or the manner of presentation in the original report. Whatever the reasons, it has been clearly illustrated that misreporting, like failed scepticism, is frequent and does provide evidence for the need for attention to these problems by the social science community.

¹² Several of the citations reviewed in this analysis provided virtually identical (word for word) descriptions of the SIPU study, the wording in their references also also being identical. For example, see and compare the descriptions offered by Harlow, Weber and Wilkins (1971:343), Public Health Publication (1971:523), and Soloman (1976:266).

IV. SUMMARY AND CONCLUSION: PREVENTING ERROR

INTRODUCTION

Despite contemporary acclaim, a detailed and careful examination of the widely and importantly cited SIPU project reveals, a shoddy study. Serious methodological problems include no adequate definition or measure of the independent variable - intensive parole supervision; a failure to adequately measure and report the dependent variable recidivism; a failure to randomly assign parolees as intended; and a lack of controls to ensure only experimentals received treatment. Additionally, the research report failed to provide information essential for a re-analysis of the data and to the study's replication. A review of fifty-five citations to the SIPU study showed only three instances of attention to these flaws. Only once has the study been interpreted as methodologically unsound.

A review of seventeen sources citing the entire four phases of the SIPU study shows that only once have the findings of the project been reported correctly. Several citations provide incorrect information in describing the study's design. Misreporting occurred regardless of whether the original research document was cited as the reference. Indeed, most of the

citations used the original document as reference.

The fact that the errors disclosed here have occurred, and are of such magnitude, suggests "organized scepticism" is not a reliable means by which error can be detected and corrected for in social science. Clearly, the problem is much worse. The errors disclosed here suggest we cannot be sure information will be reported correctly even when research findings are taken at face value. The need for additional safeguards is obvious.

ENTRENCHED ERROR

It is worth noting some of the ways in which the erroneous citations to the SIPU study have themselves been referred to in subsequent reports. Doing so allows us to see just how deeply entrenched in the literature an error can become, and thus illustrates the need to build in safeguards at the point at which original research is first slated for publication.

The consequence of at least one of the citations referred to in this thesis has been noted several times. Thus Glaser (1975-76:197) states the Lipton et al (1975) book "may have influenced more public pronouncements on correctional policy than any other writing in this century", and Halleck and Witte (1977:373) contend that "Martinson's scholarly and comprehensive work has had a powerful effect on criminal justice administrators". Similarly, Ross and McKay (1978:279) point out

that the fashionable view that treatment is a dead issue can be attributed to Lipton et al (1975) and Wilks and Martinson (1976) while Lerner (1977:211) suggests that Lipton et al (1975) has become part of the justification for proposals to abolish indeterminate sentencing and parole, and Hackler (1978:93-97) contends that the "cynical perspective" on the effectiveness of rehabilitation is exemplified in its "bible", the Lipton et al (1975) book. Finally Conrad (1978;57) concludes:

What other book in all literature has made such a progression? It was once inaccessible, therefore unread, but nevertheless an authoritative source of truth. It is now unreadable, therefore inaccessible, but still an established authority, an indispensable citation whenever correctional research is to be discussed. Its claim on opinion and belief is unshaken by criticism and likely to stand a long, if not permanent test of time.¹

¹ While Martinson's work and in particular the Lipton et al (1975) survey has been criticized (see Palmer, 1975:133-152; Empey, 1976:582-583; Adams, 1977:325; Glazer, 1978:220; Warren, 1977:359-360; Gottfredson, 1979:39-54; Gendreau and Ross, 1979:463-465), the criticism has not been, with the exception of Glaser (1978:22), directed towards a re-assessment of the methodological adequacy of the studies reviewed in the Lipton et al (1975) survey. In other words negative and positive findings were not challenged as "nugatory findings". For the most part the criticism is aimed at the conclusions offered, the manner in which the findings are presented; and to a lesser degree, the methods employed by Martinson. The criticism is such that it has probably not lessened the impact of Martinson's work, but rather only changed the direction from the notion that "nothing works" to the belief there is evidence to show "some things work". Indeed, there has recently appeared in the literature a number of articles which concentrate on showing that there is evidence to show some studies have produced positive findings (see Palmer, 1975; 133-152; Warren, 1977:355-375; Ross and McKay, 1978:279-295; Gendreau and Ross, 1979:463-489).

Several citations reviewed in this thesis have become part of the "cumulative knowledge" on criminal justice research. Thus Shover (1979:304-306) in summarizing the research on correctional effectiveness refers to the "summary statements" of Greenberg (1977), Robinson and Smith (1971) and Lipton et al (1975), while Kraus (1974:59) cites Outerbridge (1968) in support of a statement that "studies of the efficacy of 'treatment' have not demonstrated its superiority over other forms of institutional programs". Gendreau and Ross (1979:463) note the Lipton et al (1975) book and Martinson's (1974) article among the "leading examples of the numerous reports of the failure of correctional rehabilitation". Martinson's, (1974) article is also cited to support Rossi and Wright's, (1977: 39-40) statements about the inefficacy of correctional treatment. Alschuler (1979:418) and Messinger and Bittner (1979:669), writing in Criminology Review Yearbook, cite the Martinson (1974) article in support of statements about the ineffectiveness of treatment. The Martinson (1974) article is also used to support Jobson's (1977:256,269) plea to dismantle some parts of the Canadian Criminal Justice System. Finally, the Law Reform Commission of Canada (1976:117,126) refers to the "very conclusive" evidence on the efficacies of parole provided by the Waller (1974) book.

TOWARDS SAFEGUARDS

In considering how we might reduce the incident of error perpetuated in social science we have to address two separate problems; the possibility research findings and basic information about research projects will be misreported; and the possibility methodological inadequacies will go unnoticed, or at least unacknowledged. Perhaps, though, these two problems can be answered by the same set of safeguarding procedures.

There are a number of approaches to assessing of research methodology. Global peer reviews (including research reviews, reviews by journal and book editors) are the approaches the social science community has relied upon most often (McTavish et al, 1977:6). The main problems with peer reviews, however, as McTavish et al (1977:6) point out, is that "they are quite unstructured and inexplicit about the components of research methodology which are considered" and "it is difficult to tell how broadly and uniformly research standards are brought to bear in experts' judgments". Within the last fifteen years procedures for assessing research methodology have been developed in a more systematic and explicit manner (e.g. Dodd, 1966; Boalt, 1969; Bernstein and Freeman, 1975; Lin, Bingham, Heald, 1976). Even these developments however, do not incorporate checks against the "judgments" of the individual(s) using the assessment instrument. Thus, in the final analysis, they still rely on "organized scepticism", but in a more formalized way. While such

procedures are a step in the right direction, their reliability is still questionable.

Fortunately, an approach has been developed and tested recently (i.e. McTavish et al, 1977) which would, serve as an appropriate safeguard. Briefly, the approach of McTavish et al (1977) involves the independent review and rating of well documented research projects against a set of descriptive and evaluative (open-ended and structured) items by between two and five selected and experienced researchers who take the substantive merit of the project as given. Each reviewer's methodological assessment represents a "structured judgment" immediately confirmed by other "structured judgments", and therefore we know exactly what points were considered in the assessment and how accurately they were considered. The approach has been tested and found to be valid and reliable (McTavish et al, 1976:23-25).

The social science community needs to establish a standard calling for all original research slated for publication to be subjected to the approach offered by McTavish et al (1977) (or one like it), and additionally, that the assessed research be earmarked accordingly. The earmarking of research would be done using lettered and/or numerical symbols, and would be placed in the conclusion of the research report or write-up. The lettered symbol "IA", for example, could be used to signify that the methodology of the research had been assessed, and this symbol

could be followed by additional letters signifying any methodological shortcomings. Again, for example, the letters "DI", could mean "dependent variable inadequately measured", while "S", might be used to indicate "inadequate selection procedures followed". Supposing that in one study both problems were present, the earmarking following the write-up of the research would read (IA: DI,S). Once a research project was assessed and earmarked in this manner (or one like it), it could be cited with the confidence that it is as sound as it is purported to be.

This approach could also take into account the conclusions and findings of the researchers and thus provide a safeguard against misreporting. It could be established as a standard of scientific adequacy that conclusions and findings be numerically listed at the end of the research report, and earmarked to indicate the outcome of the review process (i.e. "D" for denied; "C" for confirmed, etc.). It could be further established that cited conclusions be placed in quotes, and that the earmarking also be contained within these quotes. Such a standard would allow even the reader of tertiary citations to the original research to be at least reasonably sure that the findings are exactly as they were stated in the original research document, and further, that the findings and conclusions have been scrutinized.

The proposed assessment and earmarking process has some bonus advantages aside from its capacity to reduce the creation and perpetuation of error in social science. It would make the organization of "meta evaluations"² relatively straightforward, since it would simply be a case of bringing together "reviewed conclusions". It would save each social scientist an enormous amount of time - time ordinarily spent checking the research cited as reference. Finally, the process proposed here would assist the social scientist who is not an expert in research methods, and/or data analysis and statistics. Given the increased complexity of recent research efforts, this advantage has particular significance.

²"Meta evaluation" is synonymous with the term "evaluations of evaluations". See Cook and Gruder (1978), Metaevaluation Research, Evaluation Quarterly 2:5-51; Scriven (1969), An Introduction to meta-evaluation, Educational Product Report, 2:36-38.

APPENDIX A

The criteria Lipton et al use are:

1. Insufficient data were presented to support conclusions or to permit re-analysis;
2. Data were too preliminary to make a fair evaluation;
3. Only a summary of the actual study was available and this was insufficient for a fair evaluation;
4. The findings were confounded by extraneous factors (for example, length of incarceration, idiosyncratic style of therapist) to the extent that they could not be interpreted;
5. The measures of treatment methods or outcome variables were so unreliable or invalid that findings were uninterpretable;
6. The treatment methods or outcome variables were so inadequately described that it was impossible to evaluate the study;
7. The conclusions were spurious, that is, they were unrelated to the data presented;
8. There was no indication of how the sample was selected or assigned;
9. Inadequate procedures were used to select experimental and control subjects, which made comparisons between them uninterpretable;
10. The small size of the study sample made any conclusion

hazardous;

11. Inappropriate statistical tests were used and insufficient data were presented to permit recomputation. (Lipton, et al, 1975:6-7).

APPENDIX B

Logan's criteria include:

1. There must be an adequate definition of the program or set of techniques whose effectiveness is being tested. This definition should be sufficiently operational that the components of the program can be clearly identified.
2. The technique must be capable of routinization. This does not mean that it has to be a purely mechanical activity, but it must be something that can be repeated in all its components at different times, with different subjects, by different administrators of the technique.
3. There must be some division, preferably random, of a given population of offenders into treatment and control groups, with the two groups differing as little as possible with respect to the characteristics of the subjects and their basis of selection.
4. There must be some evidence that the treatment group is in fact receiving treatment as defined, but that the control group is not.
5. There should be some "before-and-after" measurement of the behavior that is sought to be changed, and a comparison made between the two measures.
6. There must be a definition of "success" and "failure" that

is sufficiently operational to provide a valid, reliable measurement for determining the outcome of treatment.

a. This definition should be compatible with ordinary notions of what would be successful or unsuccessful outcomes of treatment.

b. There should be some follow-up or delayed measurement in the community for both the treatment and control groups.

(Logan, 1972: 378-379).

APPENDIX C

SIPU - PHASE II

Parole Violations Within 6-1/2 Months Distributed to Type of Release
Employment Assistance for Parolees Released from January through June, 1957

Type of Release Employment Assistance	Paroled	Violations	Percent
30-Man Caseloads	524	120	22.9
A. No Job	269	60	22.3
B. Family or Friend	179	36	20.1
Assisted by:			
1. None	165	33	20.0
2. Union	13	3	23.1
3. CSES	-	-	-
4. Other*	1	-	-
C. Parole Agent	76	24	31.6
Assisted by:			
1. None	36	14	38.9
2. Union	25	6	24.0
3. CSES	12	3	25.0
4. Other*	3	1	33.3
90-Man Caseloads			
A. No Job	10	2	20.0
B. Family or Friend	382	78	20.4
Assisted by:			
1. None	339	68	20.1
2. Union	42	10	23.6
3. CSES	-	-	-
4. Other*	1	-	-
C. Parole Agent	481	125	26.0
Assisted by:			
1. None	188	48	25.5
2. Union	108	24	22.2
3. CSES	141	37	26.2
4. Other*	44	16	36.4
TOTALS	1397	325	23.3
A. No Job	279	62	22.2
B. Family or Friend	561	114	20.3
Assisted by:			
1. None	504	101	20.0
2. Union	55	13	23.6
3. CSES	-	-	-
4. Other*	2	-	-
C. Parole Agent	557	149	26.7
Assisted by:			
1. None	224	62	27.7
2. Union	133	30	22.6
3. CSES	153	40	26.1
4. Other*	47	17	36.2

APPENDIX C (continued)

SIPU - PHASE II (continued)

*The "Other" category includes assistance from various charitable organizations as, Salvation Army, Volunteers of America, St. Vincents de Paul Society, etc.

Source: SIPU II:21, Department of Corrections, California, 1958.

APPENDIX D

Average Base Expectancy Score by SIPU vs. Control and
Quarter of Release*

Quarter of Release	SIPU		Control	
	Number of Subjects	Average B.E. Score	Number of Subjects	Average B.E. Score
1. July-Sep. 1957	298	15.7	677	16.3
2. Oct.-Dec. 1957	285	16.3	422	16.1
3. Jan.-Mar. 1958	168	15.9	557	15.2
4. Apr.-June 1958	95	17.2	837	14.9
5. July-Sep. 1958	35	13.3	176	14.9
6. Oct.-Dec. 1958	30	16.4	137	14.7
July 1957 - Dec. 1958	911	16.0	2,806	15.5

Source: SIPU III:24, Department of Corrections, 1962

BIBLIOGRAPHY

- Adams, R. and H.J. Vetter (1971) "Probation Caseload Size and Recidivism Rate," British Journal of Criminology, 11 (no.4) 390-393.
- Adams, S. (1967) "Some Findings from Correctional Caseload Research," Federal Probation, 31:48-57.
- Adams, S. (1974) "Measurements of Effectiveness and Efficiency in Corrections," pp. 1021-1049 in Glaser, D. (Ed.) Handbook of Criminology. Chicago: Rand McNally Publishing Co.
- Adams, S. (1976) "Evaluative Research in Corrections: Definitions, Criteria, Methods and Models," pp.235-255 in Boesen, Pand S. Grupp (Ed.) Community Based Corrections: Theory Practice, and Research. Davis Publishing Company, Inc.
- Adams, S. (1977) "Evaluating Correctional Treatments," Criminal Justice and Behaviour, 4: 323:339.
- Adams, W.O., P.M. Chandler, and M.G. Neithercutt (1971) "The San Francisco Project: A Critique," Federal Probation, pp. 45-53.
- Alschuler, A.W. (1979) "Sentencing Reforms and Prosecutorial Power: A Critique of Recent Proposals for 'Fixed' and 'Presumptive' Sentencing," Criminology Review Yearbook, 1: 416-445..
- Barnes, H.E. and N.K. Teeters (1959) New Horizons In Criminology (Third edition). Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Barnes, S.B. and R.G.A. Dolby (1970) "The Scientific Ethos: deviant viewpoint," European Journal of Sociology, 11 (no.1) 3-25..
- Bean, P. (1976) Rehabilitation and Deviance. London: Routledge and Kegan Paul.
- Boalt, G. (1969) The Sociology of Research. Carbondale: Southern Illinois University Press.
- Bernstein, I.N. (1978) "Social Control in Applied Social Science: A Study of Evaluation Researchers' Conformity to Technical Norms," Evaluation Quarterly, 7:24-47..

- Bernstein, I.N., and H.E. Freeman (1975) Academic and Entrepreneurial Research: The Consequences of Diversity in Federal Evaluation Studies. New York: Russell Sage.
- Boruch, R.F. and A.J. McSweeney, E.J. Soderstrom (1978) "Randomized Field Experiments for Program Planning, Development and Evaluation: An Illustrative Bibliography," Evaluation Quarterly, 2 (no.4) 655-695.
- Brill, R. (1978) "Implications of the Conceptual Level Matching Model for Treatment of Delinquents," Journal of Research in Crime and Delinquency, 15 (No.2) 229-246.
- Burkhart, W.R. (1976) "The Great California Parole Experiment," Federal Probation, pp. 9-15.
- Campbell, D.T., and J. Stanley (1963) Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.
- Campbell, D.T. (1979) "A Tribal Model of the Social System Vehicle Carrying Scientific Knowledge," Knowledge: Creation, Diffusion, Utilization, 1 (no.2) 181-201.
- Carey, A. (1967) "The Hawthorne Studies: A Radical Criticism," American Sociological Review, 32: 408-416.
- Carney, L.P (1977) Probation and Parole: Legal and Social Dimensions. New York: McGraw-Hill.
- Carter, R.M., and L.T. Wilkins ((1968) "Caseloads: Some Conceptual Models," pp. 391-409 in Carter, R.M., and L.T. Wilkins (Eds.) Probation, Parole and Community Corrections, (1976) 2nd ed. New York: John Wiley & Sons.
- Carter, R.M., and P.T. Takagi (1967) "Persistent Problems and Challenges in Correctional Supervision," Criminologica, 5: 36-46.
- Cavan, R.S. (1962) Criminology (third edition). New York: Thomas Crowell Company.
- Citizens' Inquiry on Parole and Criminal Justice, Inc. (1975) Prison Without Walls: Report on New York Parole. New York: Praeger Publishers.
- Clegg, R.K. (1974) Probation and Parole: Principles and Practices (Third Printing). Springfield, Illinois: Charles C. Thomas Publisher.

- Commission of Inquiry into the Non-Medical Use of Drugs - Gerald LeDain (Chairman) (1973) Final Report of the Commission of Inquiry into the Non-Medical Use of Drugs. Ottawa: Information Canada.
- Conrad, J.P. (1965) Crime and Its Correction: An International Survey of Attitudes and Practices. Los Angeles: University of California Press.
- Conrad, J.P. (1978) "Research and Development in Corrections," Federal Probation, 42:57-59.
- Cook, T.D. and D.T. Campbell (1976) "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings" pp.223-325 in Dunnette, M.D. (Ed.) Handbook of Industrial and Organizational Psychology. Chicago: Rand McNally.
- Cook, T.D., and F.L. Cook, M.M. Mark (1977) "Randomized and Quasi-Experimental Designs in Evaluation Research: An Introduction" pp.103-139 in Rutman L. (Ed.) Evaluation Research Methods: A Basic Guide. Beverly Hills: Sage publications.
- Cook, T.P. & C.H. Gruder (1978) "Metaevaluation Research," Evaluation Quarterly, 2:5-51.
- Cousineau, F.D. and S.N. Verdun-Jones (1979) "Evaluating Research into Plea Bargaining in Canada and the United States: Pitfalls Facing the Policy Makers," Canadian Journal of Criminology, 21 (No.3) 293-309.
- Dodd, S.C. (1966) "Scient-Scales for Measuring Methodology," American Behavioral Scientist, 9:3-19.
- Dressler, D. (1969) Practice and Theory of Probation and Parole (Second Edition). New York: Columbia University Press.
- Duffee, D. and R. Fitch (1976) An Introduction to Corrections: A Policy and Systems Approach. Pacific Palisades, California: Goodyear Publishing Company, Inc.
- Empey, L.T. (1976) Review of "The Effectiveness of Correctional Treatment: A Survey of Treatment Evaluation Studies, by Lipton, Martinson and Wilks," Contemporary Sociology, 5: 582-583.
- Genreau, P. and B. Ross (1979) "Effective Correctional Treatment: Bibliography for Cynics," Crime and Delinquency, October: 463-489.

- Gibbons, D.C. (1960) "Comments on the Efficacy of Criminal Treatment," Canadian Journal of Corrections, 2: 165-174.
- Gibbons, D.C. (1968) Society, Crime and Criminal Careers: An Introduction to Criminology. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Gibbons, D.C. (1973) Society, Crime and Criminal Careers: An Introduction to Criminology. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Glaser, D. (1964) The Effectiveness of a Prison and Parole System. Indianapolis: Bobbs-Merrill.
- Glaser, D. (1969a) The Effectiveness of a Prison and Parole System (Abridged ed.). Indianapolis: Bobbs-Merrill.
- Glaser, D. (1969) "Research on Probation" pp.11-46 in Cohn, A. (Ed.) Problems, Thoughts, and Processes in Criminal Justice Administration. N.P.: The Probation Management Institutes, Publication IV.
- Glaser, D. (1975-76) Review of "The Effectiveness of Correctional Treatments: A Survey of Treatment Evaluation Studies, by Lipton, Martinson and Wilks," Journal of Research in Crime and Delinquency, 12-13:179-182.
- Glaser, D. (1978) "The Counter-Productivity of Conservative Thinking About Crime," Criminology, 16 (no.2) 209-224.
- Gottfredson, D.M. (1975) "Correctional Decision-making," pp.82-91 in Gottfredson, D.M. (Ed.) Decision-making in the Criminal Justice System: Reviews and Essays. U.S. Department of Health, Education and Welfare.
- Gottfredson, M.R. (1979) "Treatment Destruction Techniques," Journal of Research in Crime and Delinquency, 16:39-54.
- Gottfredson, M.R. (1980) "Response to Lerman," Journal of Research in Crime and Delinquency, 17(no.1) 130-133.
- Greenberg, D.F. (1977) The Correctional Effects of Corrections: A Survey of Evaluations, pp.111-148 in Greenberg, D.F. (Ed.) Corrections and Punishment. Beverly Hills: Sage Publications.
- Greer, S. (1969) The Logic of Social Inquiry. Chicago: Aldine Publishing Co.

- Hackler, J. (1978) The Great Stumble Forward. Toronto: Methuen.
- Halleck, S.L., and A.D. Witte (1977). "Is Rehabilitation Dead?," Crime and Delinquency, 26:372-382.
- Harlow, E. and R. Weber, L.T. Wilkins (1971) "Community-Based Correctional Programs," Crime and Delinquency Topics: A Monography Series, National Institute of Mental Health Center for Studies of Crime and Delinquency, pp. 1-37, Government Printing Office. Reprinted in Eldefonso, E. (Ed.) Issues in Corrections: A Book of Readings, Beverly Hills: Glencoe Press, 1974.
- Havel, J. and E. Sulka (1962) Special Intensive Parole Unit: Phase Three, Technical Report. Research Division, Sacramento: California Department of Corrections.
- Havel, J. (1965) Special Intensive Parole Unit: Phase Four, "The Parole Outcome Study". Research Division, Sacramento: California Department of Corrections.
- Havel, J. (no date) Special Intensive Parole Unit: Phase Four, "High Base Expectancy Study". Research Division, Sacramento: California Department of Corrections.
- Hedley R.A. and T.C. Taveggia (1977) "Textbooks Knowledge: Some Cautionary Remarks," The American Sociologist, 12:108-116.
- Hirschi, T., and H.C. Selvin (1967) Delinquency Research: An Appraisal of Analytic Methods. New York: The Free Press.
- Institute for Scientific Information (1966-1979) Social Sciences Citation Index. Philadelphia, Pennsylvania, U.S.A.
- Jobson, K. (1976) "Dismantling the System - Our Best Alternative?," Paper presented to Symposium on Canadian Prisons, October 29-30, 1976, University of British Columbia.
- Jobson, K. (1977) "Dismantling the System," Canadian Journal of Criminology, 19:254-272.
- Kassebaum, G., D.A. Ward, and D.M. Wilner (1971) Prison Treatment and Parole Survival: An Empirical Assessment. New York: John Wiley and Sons.
- Keve, P.W. (1973) Imaginative Programming In Probation and Parole (Third Printing). Minneapolis: University of Minnesota Press, Copyright, 1967.

- Kerlinger, F.N. (1973) Foundations of Behavioral Research (2nd edition). New York: Holt, Rinehart and Winston.
- Klapmuts, N. (1973) "Community Alternatives to Prison," Crime and Delinquency Literature, 5:305-337. Reprinted in Monahan, J. (Ed.). Community Mental Health and the Criminal Justice System, pp.205-237, New York: Pergamon, 1976.)
- Kraus, J. (1974) "A Comparison of Corrective Effects of Probation and Detention on Male Juvenile Offenders," British Journal of Criminology, 14:49-52.
- Lauer, R.H. (1971) "The Scientific Legitimation of Fallacy: Neutralizing Social Change Theory," American Sociological Review, 36:881-889.
- Law Reform Commission of Canada (1976) Studies on Imprisonment. Ottawa: Supply and Services Canada.
- Lerner, M.J. (1977) "The Effectiveness of a Definite Sentence Parole Program," Criminology, 15:211-225.
- Lipton, D., R.M. Martinson, and J. Wilks (1975) The Effectiveness of Correctional Treatment. New York: Praeger.
- Lerman, P. (1980) "M. Gottfredson, Treatment Destruction Techniques. A Comment," Journal of Research in Crime and Delinquency, 17 (no.1) 126-129.
- Logan, C.H. (1972) "Evaluation Research in Crime and Delinquency: A Reappraisal," Journal of Criminal Law, Criminology and Police Science, 63:378-387.
- McCleary, R. (1975) "How Structural Variables Constrain the Parole Officer's Use of Discretionary Powers," Social Problems, 23 (no.2) 209-225.
- McTavish, D.G., and J.D. Cleary, E.E. Brent, L. Perman, K.R. Knudsen (1977) "Assessing Research Methodology: The Structure of Professional Assessments of Methodology," Sociological Methods and Research, 6 (No.1) 3-44.
- Mackie, M.M. (1973) "Arriving at 'Truth' by Definition: The Case of Stereotype Inaccuracy," Social Problems, 20: 431-447.
- Maltz, M.D. and R. McCleary (1977) "The Mathematics of Behavioral Change: Recidivism and Construct Validity," Evaluation Quarterly, 1 (No.3) 421-435.

Manual of Correctional Standards (1966) (Third Edition) (Ninth Printing, 1974). College Park, Maryland: The American Correctional Association.

Martinson, R. (1974) "What Works?: Questions and Answers About Prison Reform," The Public Interest, 35:22-54.

Merton, R.K. (1938) "Science, and the Social Order," Philosophy of Science, 5:321-337.

Merton, R.K. (1942) "Science and Technology in a Democratic Order," Journal of Legal and Political Sociology, 1:115-126.

Merton, R.K. (1957) "Priorities in Scientific Discovery," American Sociological Review, 22 (No.6) 635-659.

Merton, R.K. (1973) The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press.

Miller, E.E. (1979) "Is Parole Necessary?," pp.253-263 in Miller, E. and M.R. Montilla (Ed.) Corrections in the Community. Reston, Virginia: Reston Publishing Company.

Miller, W.B. O (1977) "Ideology and Criminal Justice Policy: Some Current Issues," pp.464-485 in Leger, R.G, and J.R. Stratton (Eds.) The Sociology of Corrections: A Book of Readings. New York: John Wiley and Sons.

Mitroff, I. (1974) "Norms and Counter-norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists," American Sociological Review. 39:595.

Morrison, D.E. and R.E. Henkel (1969) "Significant Tests Reconsidered," The American Sociologist, 4:131-140.

Memo: Lipton, Martinson and Wilks. Distributed to employees of the National Parole Service (1978).

Messinger, S.L., and E. Bittner (1979) "Part VI: Prevention, treatment and control," Criminology Review Yearbook, 1: 669-673..

Nagel, E. (1961) The Structure of Science. New York: Harcourt, Brace & World.

Nettler, G. (1978) Explaining Crime (2nd ed.). Toronto: McGraw-Hill.

O'Leary (1969) "Probation Services and the Violent Offender" pp. 175-215 in Cohn, A. (Ed.) Problems, Thoughts, and Processes in Criminal Justice Administration. N.P.: The Probation Management Institutes, Publication IV.

O'Leary, V. (1974) "Parole Administration," pp. 909-949 in Glaser, D. (Ed.) Handbook of Criminology. Chicago: Rand McNally Publishing Co.

Outerbridge, W.R. (1968) "The Tyranny of Treatment," Canadian Journal of Corrections, 10:378-387.

Outerbridge, W.R. (1970) "Re-thinking the Role of Treatment in Probation," Chitty's Law Journal, 6:189-200.

Palmer, R. (1975) "Martinson Revisited," Journal of Research in Crime and Delinquency, 12:133-152.

Public Health Service Publication (1971) Community-Based Correctional Programs: Models and Practices, Government Printing Office. Reprinted in Carter, R.M. and L.T. Wilkins (Ed.) Probation, Parole, and Community Corrections (Second Edition). New York: John Wiley, and Sons, 1976 pp.518-532.

Reckless, W.C. (1961) The Crime Problem (Third Edition). New York: Appleton - Century - Crofts, Inc.

Reckless, W.C. (1967) The Crime Problem (Fourth Edition). New York: Appleton-Century-Crofts, Inc.

Reckless, W.C. (1973) The Crime Problem (Fifth Edition). New York: Appleton-Century-Crofts, Inc.

Reynolds, L.T. (1966) "A Note on the Perpetuation of a 'Scientific' Fiction," Sociometry, 29:85-88.

Riecken H.W. and R.F. Boruch (1974) Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York: Academic Press.

Robinson, J. (1969) Technical Supplement No.2. The California Prison, Parole and Probation System. California: California Assembly.

Robison, J.O., and G. Smith (1971) "The Effectiveness of Correctional Programs," Crime and Delinquency, 17:67-80.

Roesch, R. and R. Corrado (1979) "The Policy Implications of Evaluations: Some Issues Raised by the Fishman Study of

- Rehabilitation and Diversion Services," The Journal of Criminal Law and Criminology, 70 (no.4) 530-541.
- Ross, R.R., and H.B. McKay (1978) "Behavioural Approaches to Treatment in Corrections: Requiem for a Panacea," Canadian Journal of Criminology, 20: 279-295.
- Rossi, P.H., and S.R. Wright (1977) "Evaluation Research: An Assessment of Theory, Practice and Politics," Evaluation Quarterly, 1:5-52.
- Rothman, R.A. (1971) "Textbooks and the Certification of Knowledge," The American Sociologist, 6:125-127.
- Rutman, L. (1977) "Planning an Evaluation Study" pp.15-38 in Rutman, L. (Ed.) Evaluation Research Methods: A Basic Guide. Beverly Hills: Sage.
- Savitz, L. and S. Turner, T. Dickman (1977) "The Origin of Scientific Criminology: Franz Joseph Gall as the First Criminologist," pp.41-56 in Meier, R.F. (Ed.) Theory in Criminology: Contemporary Views Beverly Hills: Sage Publications.
- Scriven, M. (1969) "An Introduction to Meta-Evaluation," Educational Product Report, 2:36-38.
- Sellin T. and M. Wolfgang (1964) The Measurement of Delinquency, New York: John Wiley and Sons.
- Shover, N. (1979) A Sociology of American Corrections. Homewood, Illinois: The Dorsey Press.
- Sjoberg, G., and R. Nett (1968) A Methodology for Social Research. New York: Harper & Row.
- Soloman, H.M. (1976) Community Corrections. Boston: Holbrook Press, Inc.
- "Special report on the Second National Criminal Justice Evaluation Workshop" (1978). Criminal Justice Newsletter, Vol. 9, Number 24, pp.4.
- State of California. "Special Intensive Parole Unit: Phase I: Fifteen Man Caseload Study," Division of Adult Paroles, W.T. Stone, Chief, Sacramento, California, November, 1956.
- State of California. "Special Intensive Parole Unit: Phase II: Thirty Man Caseload Study," Division of Adult Paroles,

Report prepared by Reimer & Warren, California, 1958.

Storer, N.W (1966) The Social System of Science. New York: Holt, Rinehart & Winston.

Suchman, E. (1967) Evaluative Research: Principles and Practice in Public Service and Social Action Programs. Beverly Hills: Sage Foundaton.

Sutherland, E.H. and D.R. Cressey (1966) Criminology (Seventh Edition). New York: Lippincott Company.

Sutherland, E.H. and D.R. Cressey (1978) Criminology (Tenth Edition). New York: Lippincott Company.

Tappan, P.W. (1960) Crime, Justice and Correction. New York: McGraw-Hill Book Company.

Waller, I. (1972) Parole for the Ex-prisoner: Carrot, Stick, or Illusion?," pp.116-134 in West, D.J. (Ed.) The Future of Parole: Commentaries on systems in Britain and U.S.A. London: Duckworth.

Waller, I. (1974) Men Released from Prison. Toronto: University of Toronto Press.

Warren, M.Q. (1977) "Correctional Treatment and Coercion," Criminal Justice and Behaviour, 4:355-375.

Williams, W. and J.W. Evans (1969) "The Politics of Evaluation: The Case of Head Start," Annals of the Academy of Political and Social Science, The Future of Corrections. N.P.: Academic of Politcal and Social Science: Book Department.

Weinstein, O. (1979) "Fraud in Science," Social Science Quarterly, 59:639-652.

Yin, R.K., E. Bingham and K.A. Heald (1976) "The Difference the Quality Makes: The Case of Literature Reviews," Sociological Methods and Research, 5:139-156.