Adult Correctional Treatment

Gerald G. Gaes
Director of Research, Federal Bureau of Prisons

Timothy J. Flanagan
Vice President for Academic Affairs, State University of New York at Brockport

Lawrence Motiuk
Director General of Research, Correctional Service Canada

Lynn Stewart
Correctional Service Canada

September, 1998

A slightly revised version of this paper will be published in Crime and Justice: A Review of Research, edited by Michael Tonry and Joan Petersilia and published by the University of Chicago Press. The opinions expressed in this paper by Gerald G. Gaes are solely those of the author and are not intended to represent the policies or practices of the Federal Bureau of Prisons or the U.S. Department of Justice.

The opinions expressed by Larry Motiuk and Lynn Stewart are solely theirs and are not intended to represent the policies or practices of the Correctional Service Canada.

Version 2, September 17, 1998

Abstract

Adult correctional treatment is effective in reducing criminal recidivism. Meta-analyses of adult and juvenile correctional interventions demonstrate that juvenile interventions are more effective than those designed for adults. Behavioral/cognitive treatments, on average, produce larger effects than other treatments. Although meta-analysis is an improvement over previous research synthesis methods, it is not without its own problems. Meta-analysis is no panacea for poorly designed and poorly executed studies, and there are still many such studies in the correctional treatment evaluation literature. Analyses of specific treatment domains also indicate effective interventions. Cognitive skills training seems successful with adult probationers and specific subgroups of offenders. Intensive, in-prison drug treatment is effective, especially when it is combined with community aftercare. Education, vocational training, and prison labor programs have modest effects on reducing criminal recidivism but they also increase positive behavior in prison. Of the four subdomains of treatment we examined closely, sex offender interventions were the most problematic. This is because the target population is heterogeneous and the treatments will probably have to be tailored to the specific offender deficits. Future theoretical development of correctional treatment principles and the use of strong inference designs that assess processes that mediate behavior change will increase our knowledge base and eventually lead to more effective adult interventions.

Version 2, September 17, 1998

Introduction

In this paper, we take on the task of reviewing the research literature on adult correctional treatment. The task is made formidable by the sheer volume of treatment domains, published, and unpublished studies. While we acknowledge the controversy that followed the publication of reviews of rehabilitation research by Lipton, Martinson, and Wilks (1975), Martinson (1974), and the National Academy of Sciences (NAS) Panel on Rehabilitative Techniques (Sechrest, White, and Brown, 1979), there is a growing body of social scientists involved in this area of research who have moved beyond the "nothing works" assertion toward a set of behavioral principles that specify the conditions under which treatment can impact post-release behavior. We view these principles more as a set of testable hypotheses rather than treatment laws. They provide a framework for future intervention research and theoretical development. We concentrate our analysis on adult rehabilitation; however, most of the intervention research so far has been targeted at juveniles. We will briefly cover this literature as well, because it serves as a foundation for understanding the theoretical and programmatic approaches of both practitioners and researchers doing adult rehabilitation research.

Our review of this literature suggests a number of conclusions. Most treatments for adult prisoners probably have modest effects. Juvenile interventions seem to have stronger effects. Behavioral and cognitive skills training seem to hold the most promise; however, we need a better classification system for these interventions so that we can be confident we are making apples-to-apples comparisons of different behavioral and cognitive treatments. Interventions in combination

(multimodal treatments) probably work better than those in isolation. If programs are to be exported, they should be exported intact, without modification, and with extensive monitoring. This is the basis for Elliot's "Blueprints" programs which we discuss in the section on meta-analysis. Cognitive skills programs are relatively inexpensive and can be seen as a foundation program – a program that can be used in conjunction with most other approaches to intervention. Intensive in-prison drug treatment programs for adults seem to work; however, they also seem to require extending treatment into the community, both during supervision, and in some cases, even after supervision has ended. Work, vocational training, and education have modest effects on adult post-release behavior; however, these programs also seem to have salutary effects on behavior while prisoners are participating in these programs during their incarceration. While there is some suggestion of effective treatment for sex offenders, of the substantive areas we reviewed, it was the most problematic and the most difficult to draw any conclusions. Although there have been significant advances in theoretical approaches to treatment, and there is a new tool -- meta-analysis -- that can be used to summarize research and deduce principles across different types of interventions, we argue that there are still too many flawed intervention studies. They are flawed in essentially the same ways as they were when the watershed reviews by Lipton et al., Martinson, and the NAS Panel were published.

A great deal of the most recent research on adult and juvenile interventions is based upon a theoretical perspective that is the culmination of work by a number of Canadian researchers. Andrews, Gendreau, Bonta, Ross and their colleagues have produced a psychological model for understanding behavioral change in offenders. In the first section of this paper, we briefly review

the tenets of this model, since it is used by many researchers in their analysis and explanation of treatment effects. In the second section of this paper, we review the results of all of the meta-analysis research that has been done thus far. We summarize this research. We also draw the readers attention to some of the problems and contradictions of treatment meta-analyses. We then briefly describe the problems we encountered when reviewing the treatment literature closely. This is not intended to be a primer of research methodology. Instead, it represents those characteristics of this research domain that we think are most problematic and present the biggest challenge to social scientists doing evaluations of adult and juvenile treatments. The next four sections of this paper cover several well-defined domains of adult intervention. We first review cognitive skills training (section III), followed by intensive in-prison drug treatment (section IV). We devote a section to educational instruction, vocational, and industrial training (section V). In the last substantive section, we focus on sex-offender treatment (section VI). There are, of course, many other rehabilitation domains such as anger management, counseling (both group and individual), psychotherapy, and values training -- just to mention a few. To do justice to all of these potential interventions would require an entire volume devoted to the prospects of rehabilitation. Since it was impossible to cover all of these types of programs individually, our approach was first to offer a more general critique of the intervention literature and, second, to use these critical principles in our assessment of specific rehabilitative subdomains. Section VII is a discussion of some remaining problems in the theoretical principles of behavior change and an explication of strong inference designs, i.e. designs that evaluate the mechanisms of change in a treatment study. We show how strong inference designs can enhance our understanding of treatment effectiveness. Finally, we summarize our analysis with some additional suggestions for

future research.

Throughout this paper, we use the terms rehabilitation, treatment, and intervention as synonyms. The definition of rehabilitation proposed by The National Academy of Sciences Panel on Rehabilitative Techniques seems particularly appropriate since it has such a broad scope. "*Rehabilitation is the result of any planned intervention that reduces an offender's further criminal activity.... The effects of maturation and the effects associated with 'fear' or 'intimidation' are excluded... ."* (original was italicized, Sechrest, White, and Brown 1979, pp. 4-5*)*.

I. The Principles of Correctional Treatment

The principles of treatment intervention that we outline below represent the collective work of Andrews et al., (1989, 1995, 1996), Andrews and Bonta (1994), Antonowicz and Ross (1994), and Gendreau and Ross (1987). Because Palmer's (1975, 1992) work, a recent book by McGuire (1995a), Lipsey's (1989, 1992, 1995) meta-analysis, and Loesel 's (1995a, 1995b, 1996) review of treatment meta-analyses are also important foundation pieces for these principles, we refer to that material as well. These principles represent the most coherent approach to treatment currently available. Even though we think all of the principles need further theoretical development and explication, most can be readily translated into testable hypotheses and meet the scientific requirement of refutability.

Criminogenic Needs. This principle mandates that intervention efforts be linked to criminogenic characteristics. Human deficits that are directly related to the propensity to commit crime are referred to as criminogenic needs. As summarized by Andrews (1995), these needs include procriminal attitudes, procriminal associates, impulsivity, weak socialization, below average verbal intelligence, a taste for risk, weak problem-solving/self-control skills, the early onset of antisocial behavior, poor parental practices, and deficits in educational, vocational, and employment skills. Static characteristics such as age, race, and gender are only important as boundary conditions that determine which interventions work for which demographic subgroup. This principle implies that we, in fact, understand individual-based attributes that determine the propensity to commit crime. It is also based upon an assumption that we have valid and reliable measurement of these needs. Most intervention research, however, rarely reports client needs except by implication. If a treatment is designed to address a deficit, then it is assumed that the need exists, rather than measuring the need prior to and after treatment.

Multimodal Programs. This principle states that <u>all</u> criminogenic deficits should be treated. If an individual has multiple deficits which, in combination, increase the propensity toward crime, these deficits must all be addressed. This principle contains an unstated assumption that researchers can assess the deficits adequately and determine the proper sequencing of treatments. Are there, for example, rudimentary cognitive and emotional deficits that must be addressed before other treatment techniques can be implemented? There is not a great deal of multimodal research examining treatment combinations, much

less their appropriate sequencing.

Responsivity. This principle states that treatment providers should match client learning styles with staff teaching styles. Programs must be tailored to the specific needs and learning styles of the clients. The unstated assumption of this principle is that we understand and can measure both learning and teaching styles so that matching can be implemented. Typically, responsivity is inferred from the success or failure of a program, rather than by specifically measuring or experimenting with different staff and client characteristics. Of all the intervention propositions, the responsivity principle seems the most obvious, yet the most tautological.

Risk Differentiation. This proposition states that higher risk clients are more likely to benefit more from treatment than are lower risk clients and that the highest level of treatment intensity should be used for the highest risk clients. This principle also suggests that there may be a small set of very high risk clients who are not amenable to treatment, or who are treatable with a great deal of difficulty. The effect of the risk principle may be modest. Although client characteristics, such as risk, do affect treatment outcomes, meta-analyses of juvenile interventions indicate that treatment seems to be more influential than client characteristics (Lipsey, 1995). In some treatment domains (see the section on cognitive skills), lower risk probationers seem to benefit more from treatment than higher risk prisoners. In the drug abuse treatment literature, the most dependent drug users benefit far less than the more moderate drug abusers from the same treatment. In one

sense, the risk principle is a restatement of the needs principle. The riskiest clients are typically those with the most needs.

Skills Oriented and Cognitive-Behavioral Treatments. This principle stipulates that treatment providers use programs that teach clients skills that allow them to understand and resist anti-social behavior. Treatment should involve effective social learning principles in order to model and shape prosocial behavior. The most comprehensive meta-analyses support this principle.

Program Implementation and Continuity of Care. This principle states that clients should be treated in well-supported programs. The best intervention will fail if there are insufficient funds, or if there is a lack of commitment from treatment staff, administrators, or support staff. This can be particularly problematic if a program is conducted within an institutional setting primarily designed for custody purposes. Community settings may be more appropriate. This principle raises the question of the suitability of prison environments for successful treatments. Some interventions may be more effective after offenders have been released to a term of supervision. This does not mean we should vitiate prison programs, since they also have a salutary effect on inmate behavior. However, the principle does imply that we need better coordination between prison and community supervision programs. The treatment initiated in an institution will be more successful if there is continued care into the community. Aftercare is more effective if it is a continuation of the type of treatment delivered in the institutional setting.

Dosage. The intervention should be comprehensive and of sufficient duration (sufficient dosage). While it is difficult to disagree with this principle, it is another matter to test optimum dosage amounts. Unlike clinical trials of medications, there are few studies of correctional treatments which examine optimum or sufficient levels. This is further complicated by the need level of the client. Thus, dosage requires fine calibration of the client as well as the treatment.

Researcher Involvement. Because of results of his meta-analysis which showed that studies that involved researchers directly in program design and development had larger effects, Lipsey (1995) argued that researchers ought to be involved in both program development and evaluation. Lipsey cautioned that researcher involvement could be interpreted as experimenter bias, i.e., the success of a program may merely be an artifact of the researcher's participation. However, he also argued that since smaller research studies yielded larger effects, the overarching principle is one of treatment integrity. Researchers can ensure the integrity of a program, if it is not delivered to a large number of people or at a large number of sites. One possibility is that both program integrity and researcher bias are taking place at the same time. Unfortunately, while both determinants produce a greater treatment effect, only treatment integrity yields an unbiased result. We argue elsewhere that program evaluations conducted by researchers testing their own theoretical proposal ought to have safeguards against possible experimenter bias. Whether or not it is possible to conduct double blind studies is moot. However, advocates of a particular intervention strategy ought to have someone else doing independent assessment

as well.

These principles form the basis of a psychological model of behavior change. While other social sciences predict behavior change on the basis of structural, cultural, or economic principles, these approaches are not easy to translate into individual-based intervention strategies. Many psychologists take the position that behavior is an interaction between predispositions and situation, where situation can be defined as structure, culture, or economic principles. One limitation of the current psychological perspective is that there may be many contexts in which behavior change, despite the best treatments, is limited by structural and cultural obstacles beyond the control of the treatment provider.

In the next section, we review the available meta-analyses of treatment research. This section gives the broadest possible overview of intervention research.

II. Meta-Analyses of Treatment Research

Any modern discussion of effective rehabilitation must include the results of meta-analyses of the treatment literature. Meta-analysis is a procedure that uses methodological and statistical tools to summarize a particular research domain (see especially Cooper and Hedges, 1994b). Prior to the development and acceptance of meta-analysis, reviews of research were typically conducted by assessing each study carefully, and then tallying those studies that supported a theoretical perspective against those that did not. The major weakness of this 'vote counting' technique was

that it did not allow one to mathematically summarize studies. Furthermore, studies conducted with large samples were typically weighted as if they were just as important as studies with small samples. Thus, a treatment effect observed on 10 clients was considered to be just as important as a treatment effect observed on 1,000 clients. In contrast, meta-analysis is based on a common metric, called an effect size, that is standardized and is used to represent the results from each study. There has been substantial development in the statistical and methodological theory that supports the conditions under which effect sizes can be analyzed.

To include a study in a meta-analysis, there must be sufficient information to calculate the effect size. The researcher assigns a value to the impact of the independent variable on the outcome being studied. For example, one common metric that is often chosen is a correlation. If treatment has an effect in a given study, then that effect can be summarized by the correlation between the treatment and the study outcome. If a study has a very large sample size, then that correlation can be given a higher weight than studies having smaller samples. One can average all of the correlations and discuss the average effect of all of the studies, or one can discuss the distribution. In the latter case, the researcher can discuss the studies with the most typical, the largest, or the smallest effect.

Although the main purpose of a meta-analysis is to calculate the effect size for each study representing the treatment effect under consideration, the technique has more value when other measures are coded that allow the researcher to interpret the precise meaning of the results of the research synthesis. Measures of study methodology, treatment type, characteristics of the sample,

and other variables can help the researcher set the boundary conditions of a treatment. These measures have been called moderator variables. They are very important in suggesting not only the limiting conditions of a treatment, but future directions of possible theoretical interest. In much the same way as one would conduct an analysis on a sample of subjects, a meta-analysis can be conducted on a sample of studies. Instead of variables representing subject characteristics, a meta-analysis includes moderator variables that indicate specific conditions that are related to the variation in treatment effects. This approach provides a much more rigorous and systematic analysis of effects, their causes, and their limiting conditions.

Although meta-analysis has gained wide appeal and is used extensively in the fields of medicine, education, and social science, certain precautions are necessary. Although meta-analysis does improve the rigor of research reviews, the consumer must evaluate the results of such studies under the same scientific standards that would be used with any primary research or secondary summarization of that research (Hall, Tickle, Degnen, Rosenthal, and Moesteller, 1994). The same questions apply. Do the results make theoretical sense? Is there consistency across different meta-analyses, especially meta-analyses of the same set of studies? Are there reasons to suspect that there is some fundamental bias broadly underlying the results of a particular research domain?

To the unfamiliar consumer of meta-analysis, it is perplexing how a statistical technique can be used to synthesize the "truth" from studies that individually may have serious methodological flaws. Researchers conducting the best meta-analyses try to minimize these problems by selecting

the better designed studies and by coding methods variables that indicate when the conclusions of a meta-analysis are more likely artifactual than real.

In the next subsection of this paper, we review the meta-analytic studies of correctional treatment. We describe what they imply about effective juvenile and adult intervention. For the reader interested in what we see as the more troublesome conceptual and methodological problems with this literature, we have addressed those issues in Appendix I.

A. Current Meta-analyses

Lipton, Pearson, Cleland, and Yee (1998) have identified 28 reviews of the intervention literature. Of these, about half were meta-analyses. These studies are depicted in Table 1. Loesel (1995a) has published a review of meta-analyses of correctional treatment. At the time of his publication, he summarized 13 meta-analyses. Some of the studies focus on a specific intervention subset. For example, Tobler's (1986) meta-analysis was on youth drug prevention programs. Mayer, Gensheimer, Davidson, and Gottshalk (1986) examined juvenile social learning treatment. Garrett (1985) examined the effect of residential treatment for adjudicated juveniles.

Table 1 also lists the number of studies analyzed and whether the study population was juvenile or adult. Confirming Loesel's assessment, most of the meta-analyses were based on the juvenile intervention literature. Mark Lipsey's (1992, 1995) analysis of juvenile interventions supersedes, and is inclusive of, almost all of the other meta-analyses of juvenile interventions. He collected

juvenile intervention studies, published and unpublished, from English-speaking countries, produced between 1950 to 1991. He analyzed 443 of the most rigorously designed studies that used recidivism as an outcome. Not only does Lipsey's analysis cover most of the other studies included in other meta-analyses of juvenile treatments, it also has been the most thorough assessment of moderator variables.

The CDATE project (Lipton, Pearson, Cleland, and Yee, 1998) will be the most comprehensive review and research synthesis of correctional treatment ever undertaken, The project will include juvenile and adult intervention studies, published and unpublished, from any source country, produced between 1968 to 1997. Of the 1,500 studies included in the CDATE project thus far, 900 used recidivism as an outcome. The preliminary findings reported by Lipton et al., have been based on 226 juvenile studies and 261 adult studies.

B. Meta-Analytic Findings From the Treatment Literature

Although his meta-analysis will be superseded by the CDATE project, by far, the most comprehensive published meta-analysis of treatment has been Lipsey's analysis of over 400 studies representing juvenile delinquency interventions. Lipsey found that after controlling for differences in study methodology, he was able to find a substantial effect due to treatment variations. Treatment in public facilities, custodial institutions, and the juvenile justice system produced smaller effects than treatment in community settings. Treatment that was behavioral, skill-oriented, or multi-modal (utilizing more than one treatment approach) was associated with larger

effect sizes than treatment that was based on deterrence, family counseling, group counseling, or individual counseling. Because the treatment effect depended on whether the treatment was delivered in a juvenile justice or community setting, these conclusions have to be qualified accordingly. For example, employment training had the highest average effect size when it was delivered in the juvenile justice setting. When it was delivered in non-juvenile justice settings, it actually had a negative effect size, indicating that the intervention increased delinquency.

At face value, Lipsey's results are very encouraging, especially for juvenile justice interventions. By developing a large set of methods variables, Lipsey was able to show how variation in treatment effects are indeed due to variations in the quality of the research methods. He was also able to show that despite this method variation, juvenile correctional treatments produced quite dramatic effects.

Because the CDATE project includes both juvenile and adult interventions, we can begin to see emerging patterns in the differences between these two populations. Preliminary data from the CDATE project (Lipton, Pearson, Cleland, and Yee, 1998) indicates cognitive/behavioral treatment, on average, produces larger effects than interventions characterized as punishment, intensive community supervision, educational training, substance abuse, or group counseling. The CDATE data also show that with a few exceptions, randomized designs produced smaller effects than non-random designs. This again supports the proposition that methods variables must be analyzed to do a comprehensive meta-analysis. The CDATE project has also shown that juvenile interventions are typically more effective than adult interventions. However, some of these

comparisons rely on very small samples of adult studies. While moderator variables such as type of treatment and type of target population indicate the boundaries of treatment effectiveness, most meta-analyses have been exploratory and unguided by any particular theoretical perspective. One major exception, however, has been the work of Andrews, Zinger, Hoge, Bonta, Gendreau, and Cullen (1990).

C. Comparability of Results from Treatment Meta-analyses

An important question to be answered is whether or not the meta-analyses conducted thus far have produced similar results. Unfortunately, it is difficult to compare the meta-analytic studies because they do not always overlap in their publication universe, i.e. in their inclusion or exclusion of studies. In the case of the juvenile literature, Lipsey's study subsumes all or most of the other meta-analytic studies of juvenile treatment. Furthermore, the meta-analyses vary in their methodological, theoretical, and statistical rigor. Simply stated, some meta-analyses are better than others. In our opinion, it is improper to summarize these studies as if each meta-analysis should be given equal consideration.

We are most interested, of course, in whether certain juvenile and adult interventions are better than others. If we compare the two most comprehensive assessments, Lipsey's analysis and preliminary results from the Lipton et. al., CDATE project, we can point to some tentative conclusions. In their preliminary analysis of the CDATE data, Lipton et al., identified the following types of interventions: punishment, community supervision, education programs,

cognitive behavioral/social learning programs, substances abuse, and group counseling. Lipsey used the following treatment categories: employment, multi-modal (more than one treatment type), behavioral, skill oriented, institutional, community residential, individual counseling, group counseling, family counseling, vocational training, deterrence, and several kinds of probation/parole interventions.

Unfortunately, thus far, only group counseling and behavioral programs are categories of treatment represented in both meta-analyses. Lipsey's study indicated that the average residualized effect size (after controlling for methods variables) for group counseling conducted within the juvenile justice system was .07. The effect size was .18 in a non-juvenile justice setting. The CDATE results, so far, indicate group counseling has had little or no effect on recidivism. Lipton et. al., found an average effect size of .01 for juveniles participating in non-random studies and an average effect size of .00 for juveniles participating in random assignment studies. It is not clear why the CDATE and Lipsey's results are so discrepant with respect to group counseling. However, there is more consistency in the results of the two meta-analyses when comparing studies defined as cognitive behavioral or social learning interventions. Lipsey found juvenile behavioral studies to have consistently high effect sizes, whether the intervention occurred in a juvenile justice setting (effect size .25) or non-juvenile justice setting (effect size .20). The CDATE project indicated average effect sizes for cognitive behavioral and social learning programs of .18 for juveniles in non-random assignment studies and .16 for juveniles in random assignment studies. Despite this apparent similarity in findings, it is difficult to know whether researchers are using consistent definitions in their categorization of studies, and thus, whether the

consistency or inconsistency in effect sizes across meta-analyses are apparent or real. To examine this question empirically, a public domain data set composed of all of the studies and the variables that have been coded on each study would allow other researchers to compare and contrast these studies as well as conduct their own meta-analysis that builds on the previous research. The CDATE project is supposed to result in such a public domain data set.

The CDATE analysis also shows that for adults, while effects of treatment have been less pronounced, there seem to be modest effects of education, cognitive skills/social learning programs, substance abuse treatment, and group counseling. If we assume that the random assignment studies are more methodologically rigorous, and therefore, more likely to represent the true effect size of a treatment, then the average effect sizes for the adult treatments range between .03 to .06. These are very modest effects.

D. The Substantive Meaning of Effect Sizes

One common way of representing the substantive or practical impact of an intervention is to convert an effect size to a binomial effect size display, or BESD (Rosenthal and Rubin, 1982). The BESD represents the change in success (or failure) rate attributable to a treatment procedure. A correlation coefficient of .10 represents a BESD of .10. Thus, an experimental manipulation having a BESD of .10 would increase the success rate from 45 to 55 percent. A BESD of .05 would increase the success rate from, 47.5 percent to 52.5 percent, and a BESD of .20 represents an increase in the success rate from 40 to 60 percent.

Based on the juvenile meta-analyses, and especially Lipsey's study, a conservative average effect size for juvenile interventions is .10. Loesel (1995a) has also concluded that .10 is a conservative, yet reasonable effect size for juvenile treatments. This translates into a reduction in recidivism from 55 to 45 percent. The magnitude of this effect is substantively quite large.
The more modest effect sizes observed in the CDATE adult interventions would represent an average reduction in recidivism from about 52.5 percent to 47.5 percent.

E. The Blueprint Paradigm and the "Preventing Crime" Report

In this subsection, we take a brief digression from meta-analysis to consider two alternative approaches to assessing interventions. The "blueprints" approach uses a set of criteria that must be met for an intervention to be considered a model program. Similarly, researchers at the University of Maryland developed a scientific scale to rate programs and used that to specify model programs.

While meta-analysis could be construed as a tool for investigating the principles of intervention, a pragmatic approach is to export exemplary programs with proven track records. Elliot (1997) has attempted to do just that in a series entitled "Blueprints for Violence Prevention." Using a stringent set of program effectiveness criteria, Elliot and his colleagues examined more than 400 delinquency, drug, and violence prevention programs for juveniles. To date, 10 programs have met the selection criteria. These programs had a strong research design, showed evidence of significant deterrence effects, produced effects that were sustained well beyond the end of the

intervention, and produced effective results at more than one site. Elliot (1997) also looked for evidence of a strong inference design. Although he used different terminology than we have adopted, Elliot expressed surprise that many programs had not collected the necessary data to do a strong inference analysis.

The blueprints approach is atheoretical in the sense that a program must meet primarily methodological rather than theoretical standards. In fact, Elliot emphasizes that blueprint programs are to be adopted without modification. This is because there is no evidence to identify which features of a program are those that make it work. This approach may be a practical solution to treatment in the absence of good theory and well-supported principles. However, it seems to us that a more theoretically driven approach will eventually result in a more generalizable set of principles and programs. Despite the problems of the rehabilitation and treatment meta-analysis literature, we view the blueprints technique as a stopgap, yet sound practical solution in the short run. In the long run, correctional treatment effectiveness must rely on the plodding, conservative, trial-and-error nature of original research.

Another approach to the synthesis of rehabilitation research was a byproduct of an even broader analysis of programs that could reduce crime in different institutional settings such as communities, families, labor markets, and places such as businesses. The analysis also covered interventions by police and criminal justice agencies once a defendant had been arrested or convicted. The study was conducted by Sherman, Gottfredson, MacKenzie, Eck, Reuter, and Bushway (1998) for the National Institute of Justice and is entitled “Preventing Crime: What

Works, What Doesn't, What's Promising." Similar to Elliot's blueprint paradigm, Sherman et. al., developed a set of criteria to rate different program interventions. They developed a scale called the "Maryland Scale of Scientific Methods." The scale has a rating of 1 (lowest rating) to 5 (highest rating) and is based upon the methodological rigor of the study. Generally, randomized designs achieved the highest rating. Studies which indicated a relationship between a program and a crime but had few controls received the lowest rating. Doris Layton MacKenzie wrote the chapter on criminal justice institutions (chapter 9). In her assessment of rehabilitation programs, in addition to the scientific methods scale, MacKenzie used published reviews of the literature and meta-analyses of the adult and juvenile programs. She concluded that in the area of adult and juvenile correctional treatment, two types of programs demonstrated consistent effectiveness, risk-focused interventions and prison residential drug treatment. While we generally agree with the "Preventing Crime" conclusions on correctional treatment, a close examination of even these two types of interventions suggests some important qualifications. In later sections of this report, we analyze the prison residential drug treatment literature and return many times to the risk principle and focused interventions.

F. Summary of Meta-analyses

In a research area as broad as the treatment domain, meta-analysis has given us perspective where there existed primarily a great deal of confusion. Has it given us the correct perspective? From one point of view, meta-analysis has been used to draw sweeping inferences from a great many studies, many of which were poorly designed. Even if meta-analysis can be used this way, to

separate the wheat from the chaff, ultimately, meta-analysis should suggest a definitive set of studies. Those studies must be conducted to conclusively confirm or disconfirm the implications of the meta-analyses. If, through meta-analysis, we learn that interventions must be multi-modal, behavior oriented, and tailored to the client, to test these assumptions, we should design and implement a series of studies. These studies should incorporate different staff, multiple treatment sites, and several evaluators in the most rigorous designs possible, having sufficient power and sufficient internal and external validity to justify drawing conclusions from the findings. Meta-analysis should not be an end unto itself. Cooper and Hedges (1994a) make this same point succinctly. "A research synthesis should never be considered a replacement for new primary research. Primary research and its synthesis are complementary parts of a common process...(p.524)"

The next four sections of this paper are devoted to a summary of specific areas of adult correctional interventions: cognitive skills training, intensive drug treatment, education and work programs, and sex offender treatment. We make specific criticisms of studies where it is warranted. However, there are several problems worth describing apart from the specific literature, since these are problems that are endemic to treatment evaluation studies. Too few studies provide any detail on the treatment being delivered. Nor do most evaluation studies report on whether there has been any attempt to monitor the quality of the intervention. Few studies use what has been described as a strong inference design (Platt, 1966). A strong inference design is one in which the evaluation measures the level of the offender's need or deficit prior to, and after the treatment. In such a design, in addition to observing a treatment effect based on differences in

the mean level between treatment and comparison subjects, the researcher also analyzes whether treatment reduced the client's needs. A strong inference occurs when there is both a treatment effect on the group and it can be demonstrated that the reduction in a client's deficits was related to the client's outcome. For example, cognitive skills could be measured prior to and after cognitive skills training. If the offenders in the treatment group were less likely to recidivate after their treatment, that would be taken as evidence of a treatment effect. In addition, if those inmates with the largest increase in cognitive skills ability were the least likely to recidivate, we could make a much stronger inference that the cognitive skills training caused the observed effect. One other persistent problem in this literature is the one caused by subject selection and attrition. We have read too many studies in which inmates were allowed to self select into treatment, or were selected by another agent, or dropped out of treatment. In this latter case, either there was no attempt to measure the outcomes of dropouts, or their outcomes were analyzed as if they could be considered independently from clients who completed treatment. In each of these cases, the selection process may have resulted in a biased outcome. An exposition of this problem is represented in Gaes (1997, 1998) in the context of in-prison drug treatment. In the following review of the four treatment areas, we point out specific instances of each of these problems.

III. Cognitive Skills Training

In the last 15 years, cognitive-behavioral interventions have been identified as the treatment approach most often associated with reductions in offender recidivism (Lipton, 1998; Gendreau and Ross, 1979; 1987; Izzo and Ross, 1990). Researchers have come to agree on the

effectiveness of this treatment approach based on meta-analyses, qualitative analyses, and reviews of the component elements of successful programs. Gendreau and Andrews' review of meta-analytic studies of correctional treatment led them to conclude that "types of intervention should be behavioral in nature with emphasis on cognitive and skill building techniques (1990:182)." More recently, the CDATE meta-analysis conducted by Lipton et. al., confirmed that the cognitive approach produces reductions in criminal recidivism. After closely reviewing the cognitive skills research, we also agree that the program can be effective. It appears to be less effective with the highest risk participants and with specific subgroups such as younger adults.

A number of cognitive skills programs are now delivered in correctional settings. However, no one program seems to have been so widely adopted as the Cognitive Thinking Skills program developed by Robert Ross and Elizabeth Fabiano. This has become a core program in the federal Canadian correctional system. It has been implemented in the United States, Europe, Australia, and New Zealand, and throughout the British Prison system and the Probation Service in the United Kingdom. Cognitive Thinking Skills meets all the criteria of effective correctional programs. It is cognitive and behavioral in design. It addresses needs that are empirically associated with criminal behavior. It is multi-model in design, providing a number of effective behavioral and skill based treatment techniques to address a range of relevant targets. It is longer in duration than most thinking skills program, thus, its dosage or strength is higher than many similar treatment programs. Finally, in screening offenders who have significant thinking deficits, the program adheres to the risk principle in that higher risk offenders are more likely to have these thinking deficits. When properly implemented, therefore, one would expect that the delivery of

the program would provide a test of rehabilitation advocates' contention that appropriate cognitive behavioral programs can significantly reduce recidivism. This section briefly describes this program, its development, and implementation. In subsequent sections, we review the literature of cognitive skills program evaluations.

The Cognitive Thinking Skills Program, also known as the Reasoning and Rehabilitation Program, was developed by a systematic process that began with a review of all controlled evaluations of correctional programs published between 1973 and 1978 that reduced criminal recidivism (Gendreau and Ross, 1979). Ross and Fabiano (1985) identified 100 evaluations of effective programs and found that all of them applied techniques designed to target offenders' thinking. Ross and Fabiano then engaged in a literature search to identify which of the offenders' specific cognitive deficits were linked to criminality. Among offenders who demonstrated a repetitive pattern of criminal behavior, the authors identified problems with impulsivity associated with poor verbal self-regulation, impairment in means-end reasoning, a concrete thinking style that impinges on the ability to appreciate the thoughts and feelings of others, conceptual rigidity that inclines them to a repetitive pattern of self defeating behavior, poor interpersonal problem solving skills, egocentricity, poor critical reasoning, and a selfish perspective that tends to make them focus only on how their actions affect themselves instead of considering the effects of their actions on others.

Next, Ross and Fabiano scanned the literature for interventions that successfully addressed each of these deficit areas. Those interventions became the core components of the Cognitive Thinking

Skills program. They found that impulsivity could be reduced by teaching consequential thinking. Fatalistic thinking could be reduced by teaching offenders meta-cognitive skills which enabled them to assess the role their thinking has in influencing their actions. Antisocial behavior could be diminished by teaching offenders to replace these behaviors with prosocial ones. Rigid thinking could be minimized by teaching offenders creative thinking skills to provide them with prosocial alternatives in responding to interpersonal problems. Illogical thinking could be modified by critical reasoning skills. Egocentrism could be overcome by teaching offenders social perspective taking and values enhancement. Finally, social adjustment could be improved by training offenders in self control techniques.

Each of the skill training modules described above is delivered over several sessions with considerable overlap in material designed to provide adequate opportunity to overlearn the skills. All of the techniques were selected from programs that had already demonstrated a degree of effectiveness in developing the target skills. The authors acknowledge that their final program, consisting of 35 2-hour sessions, is an amalgam of content and techniques borrowed from a number of sources. The program is delivered to groups of four to ten offenders, two to four times per week. The trainers' manual is highly organized and scripted to maximize the standardization of the program. A key to the successful delivery of the program has been the selection of a variety of training techniques that create an enjoyable classroom experience for the participants. The program avoids a didactic presentation of material. Rather, the trainers or coaches, as they are called, use role playing, video-taped feedback, modeling, group discussion, games, and practical homework review to teach the skills. Another important component of the

program's success is the careful selection, training, and monitoring of the non-professional staff who deliver the program.

A. Outcome Studies

Despite its availability as a program that has been relatively stable now for over 10 years, and its delivery in dozens of sites throughout the world, there are surprisingly few controlled outcome studies of the Cognitive Thinking Skills program. The following section reviews the few published studies, some internal reports, and some preliminary findings from sites where the program has been more recently implemented.

The initial evaluation study (Ross, Fabiano, and Ewles, 1988) was conducted by the developers who delivered the program to a group of high risk adult probationers in Ontario, Canada. The probationers were selected based on their high risk rating on the Level of Supervision Inventory (Andrews, 1989), an instrument empirically well validated as a measure of recidivism risk. Probationers were randomly assigned to one of three conditions, the Cognitive Skills program, a Life Skills program, and Regular Probation without any program intervention. The probation officers taught both the Cognitive Skills and the Life Skills program, and also supervised the offenders. Recidivism was calculated based on official convictions for new offenses after a period of 9 months. The results indicated that the Cognitive Skills program was significantly more effective in reducing recidivism in the short term than either Regular Probation or the Life Skills program. Recidivism for the Cognitive Skills group was 18.1 percent (4/22). For the Regular

Probation group, it was 69.5 percent (16/23). For the Life Skills group, it was 47.5 percent (8/17). The differences between the three groups were statistically significant. The authors also evaluated the program's effect based on the percentage of offenders who received custody sentences. While none of the Cognitive Skills group received a custody sentence, 30 percent (7/23) of the Regular Probation group were reincarcerated and 11 percent (2/17) of the Life Skills group received a sentence of imprisonment. Unfortunately, although the design did include a pre- and post-test battery, the offenders were not available to complete the post-testing.

In the late 1980's, the Cognitive Thinking Skills program was piloted in Correctional Service Canada (CSC), Canada's federal correctional service. CSC is responsible for the custody and supervision of adult offenders sentenced to two or more years of imprisonment. The initial pilot evaluation involved 47 incarcerated offenders who were selected because of deficits in areas targeted by the program. The control group was composed of 26 prisoners who met the program criteria, who had agreed to treatment, and were on a waiting list (Robinson, Grossman and Porporino, 1991). Analysis of the profiles of the two groups indicated that the program criteria generally screened in higher risk offenders. The control and treatment groups were not significantly different on important characteristics such as criminal history, risk level, and demographic measures. The correctional staff who had been selected as trainers were themselves trained in the content and delivery of the program by one of the program's developers, Liz Fabiano. Offender recidivism was measured after the inmates had been released to the community. At the time of the initial follow up, ten offenders in the study had still not been released from custody. The remaining subjects were followed for an average period of 18 months. Although

the numbers were too small to detect a significant difference, the recidivism rate differences between the two groups imply that a treatment effect may have occurred had a larger sample been selected. Twenty percent (8/40) of the treatment group were reconvicted, while 30.4 percent (7/23) of the control group were reconvicted within the follow-up period. Robinson et. al, (1991) also measured readmissions without a new conviction. When the data were analyzed using readmissions without a new conviction as the outcome, the effect was in the opposite direction. Twenty-five percent (10/40) of the treatment group, but only 21.7 percent (5/23) of the controls were readmitted without new convictions.

The authors also examined the rates of reconviction based on risk levels that were calculated using the Statistical Information on Recidivism (SIR) scale (Nuffield, 1982). Based on the SIR, after 2.5 years, the average probability of reconviction for offenders with risk scores similar to the subjects in the treatment and control groups is .52. While the actual probability for the treatment group was much lower at .22., the follow up period was, on average, one year shorter than the follow up period used to make the SIR predictions. Robinson et. al., reasoned, however, that most offenders recidivate after one year of release, so that the rates of return should not be altered significantly after one year. Since the control group also had a substantially lower recidivism rate than that predicted by the SIR risk score, it is probably improper to try to draw any comparison between either the study or control group and a SIR-imputed base rate.

A second pilot study (Porporino, Fabiano, and Robinson, 1991) followed a sample of 54 offenders, (42 program completers and 12 dropouts) who had been released to the community for

an average follow up period of 12 months. Although there was no comparison group, the reoffending rate for the treatment group, 4.7 percent, was extremely low. As discussed above, based on the SIR, the expected base rates after 2.5 years would be 52 percent reconvicted. The reconviction rate of the dropout group, however, was 33.3 percent, suggesting again that the offenders who remained in the treatment group may have been among those most predisposed to do well on release.

Following the success of the pilots, the program was implemented nationally. National implementation meant the establishment of a training and monitoring infrastructure that ensured program integrity. Staff awareness training was also held to establish the support required for a well integrated correctional program. An extensive assessment battery was implemented to measure pre- and post-program status on a variety of attitude and cognitive skills scales. A great deal of demographic information was also collected on the offenders. An early report (Robinson, Grossman, and Porporino, 1991) assessed changes on a sample of 200 treatment participants and 54 comparison subjects who were on a waiting list to participate (wait list controls). The treatment group improved on all the attitude tests, as well as the measures of impulsiveness and empathy. The changes were significantly better than those of the wait list controls on 4 of the 9 scales. On the cognitive abilities scales, the treatment group improved significantly on all the relevant scales, and improved significantly more that the wait list controls. This report, however, did not provide recidivism data that would have permitted a conclusion that the gains on the program goals were significantly related to reductions in recidivism.

To date, the largest scale evaluation of the Cognitive Skills program that has been completed was based on a sample of 2,125 released Canadian federal offenders -- 1,444 program completers, 302 program drop-outs, and 379 wait list controls (Robinson, 1995). Offenders were randomly assigned to the treatment or wait list group in the first years of implementation of the program in CSC (thereafter, program implementation was expanded allowing for the treatment of all referred offenders). Recidivism outcome data were collected after inmates had been released to the community for a minimum of one year. Overall, the results indicated a small significant difference favoring the treatment group. Forty-four and a half percent of the program completers were readmitted during the first year in the community compared to 50.1 percent of the wait list controls. When the outcome criteria was reconviction, there was again a small significant difference between the groups. For the treatment group, 19.7 percent received a new conviction, while 24.8 percent of the control group received a new conviction. In both cases, the worst outcomes were for the dropouts, whose readmission and reconviction rates were 58.2 percent and 28.8 percent respectively. Comparison of the wait list controls and the treatment groups demonstrated that the wait list was composed of significantly more property offenders. To control for these differences in conviction offense, an analysis of covariance was conducted. Regardless of whether outcome was measured as readmission or reconviction, the analysis did not demonstrate any statistically significant differences between the treated and untreated offenders.

Although the overall results were encouraging, Robinson did a further set of analyses to examine if there were particular subgroups that may have benefitted more than others from the program. This kind of post hoc analysis should only be considered suggestive. Robinson's analysis of

subgroups was based on offense profiles, risk and need levels, and treatment settings. These analyses demonstrated, contrary to the risk principle, that the program was significantly more helpful for the lower rather than higher risk offenders. However, despite their designation as a "lower" risk group, these offenders had an average one year readmission rate of 36.2 percent, which is quite high. The reanalysis also showed that the medium and high needs offenders, as well as violent, sex, and drug offenders were more likely to benefit from treatment than the offenders with fewer deficits and those offenders who were property offenders.

The largest treatment effect was found for those offenders who received the program in the community. The study identified 131 subjects who completed their program in the community. However, in the absence of a community based control group, Robinson compared the community treatment group to the control group used for the institutional sample. Robinson used statistical procedures to control for the following offense categories: non-violent property offenses, robbery offenses, and violent offense. The results demonstrated a significant treatment effect on both readmissions and reconvictions. He found that 30.5 percent of the community treatment completers compared to 50.1 percent of the institutional control group were readmitted. He also found that 8.4 percent of the community treatment group as compared to 24.8 percent of the institutional control group were reconvicted during the follow-up period. These latter data appear to confirm the evidence from previous reviews of correctional treatment which emphasize greater success of a program delivered in community rather than institutional settings (Andrews, et al, 1990; Izzo and Ross, 1990: Loesel, 1995a).

The conclusions of this study are compromised by several design problems. The outcomes of the treatment completers and dropouts were analyzed separately. As we have argued in several other parts of this paper, this has the effect of biasing the results in favor of finding a treatment effect. The analysis of the community treatment sample had additional problems. Although Robinson applied statistical techniques to control for offense background, the wait list control group was composed of institutional subjects who were likely to be higher risk offenders, since they have not yet been vetted by a parole board that would have been less likely to release the highest risk offenders. Robinson's analysis of covariance apparently did not include overall risk level as a control. Thus, the comparison between the community treatment group and the institutional control group may not have been an adequate test of the effectiveness of the program. Secondly, half of the community treatment population came from one Region. In the absence of evidence that the readmission and reconviction base rates from this Region were comparable to the other Regions, it is not possible to dismiss the possibility that the treatment effect may actually reflect a generally lower risk treatment sample that returns to generally less criminogenic communities. Finally, the dropout rate in the community treatment sample was much higher than that of the institutions (30.6 percent of the community sample compared to 14.2 percent of the institutional sample). This would have the effect of biasing the community based outcome even more than the institutional outcomes. The high drop out rate was likely to have distilled the most motivated and stable offenders who remained in the treatment group.

In summary then, the largest scale outcome study of the Cognitive Skills program, to date, provided some modest evidence of the effectiveness of the program in reducing returns to custody

and reconvictions for incarcerated adults. The study did suggest the direction for future research. The most resistant offenders to the impact of this intervention were property and non-violent robbery offenders and offenders under age 25. The report does not provide the results of the pre- and post-evaluations. However, Robinson (personal communication) indicated, that although the treatment completers improved on almost all of the measures of program objectives, these improvements were not related to their reoffending. On the basis of this result, the Research Branch of the Correctional Service of Canada has redesigned the evaluation framework, developing measures that more directly assessed the program content. To date, there have been no outcome studies assessing the new evaluation framework. This effort is consistent with our argument about developing strong inference designs in program evaluations.

Outside of the Correctional Service of Canada, there have been a number of studies that have evaluated the effectiveness of cognitive skills training on offenders. The best controlled of these is the study conducted for the Colorado Judicial Department on the specialized drug offender program (SDOP) (Johnson and Hunter, 1992). Prior to their data collection, the researchers developed a comprehensive measure designed to assess the progress of clients against the program's objectives. The objectives included the reduction of drug use and other factors associated with criminal behavior, as well as improvement in skills and attitudes conducive to a more productive prosocial lifestyle. All the factors were selected based on their relevance to the goals of the program and their association with subsequent drug use or unlawful behavior.

The study randomly assigned 134 offenders with significant problems with addictions to one of

three groups: the SDOP with the Cognitive Skills group participation, the SDOP without the Cognitive Skills program, and Regular Probation. Johnson and Hunter obtained complete pre-test batteries from 124 of the clients, while only 80 completed the post-test. After an average follow up period of 8 months, cognitive skills training did seem to reduce the probability of revocation. The revocation rates were 41.7 percent (15/36) for the Regular Probation group, 29.4 percent (15/51) for the SDOP/ Non Cognitive group, and 25.5 percent (12/47) for the SDOP/Cognitive group. In this study, the risk principle was confirmed. The treatment effect was more significant with the high risk clients and for those with very high scores on the Addictions Severity Index (ASI). The high risk clients who participated in the SDOP, either with or without the Cognitive program, had a revocation rate of 35 percent, while the high risk clients in the Regular Probation group had a revocation rate of 75 percent. Those clients with high addiction scores who participated in either SDOP treatment group had a revocation rate 60 percent lower than highly addicted clients in the Regular Probation group. In this short term follow-up, the SDOP/ Non Cognitive Program option appeared to be most effective with the younger, more disturbed clients. The SDOP/Cognitive option was more effective with the older clients (over age 30) and those with severe substance abuse scores. Diagnostic and survey measures suggest that the cognitive program intervention was not effective with offenders scoring higher on measures associated with psychiatric problems and sociopathy (high normlessness, low empathy, and low belief that criminal behavior is wrong). Post- test results on the measures described above that assessed the extent to which the program met its identified goals indicated, that for the majority of needs, the success rates for clients in the SDOP/Cognitive group were higher than success rates in the SDOP/Non Cognitive group. On 16 of the 18 dimensions, the success rates for the two SDOP

treatment groups were better than success rates for the Regular Probation group.

In a later publication, Johnson and Hunter (1995) reported on an extension of the study. Examining the survivors of the initial study, 35 percent (8/23) of the Regular Probation, 30 percent (10/33) of the SDOP/Non Cognitive group, and 19 percent (6/32) of the SDOP Cognitive group were revoked at the end of another year. According to the researchers, this demonstrated that the treatment effect for the SDOP/Cognitive group did not diminish over a longer term. Because this was an analysis of survivors, these results could be attributed to characteristics of the remaining clients, and not the long term effect of treatment. A better approach to the survivability of offenders could be addressed with a well-controlled event history analysis that would indicate the hazard of revocation over time.

In the United Kingdom, the Cognitive Skills program is known as the Reasoning and Rehabilitation program. In the community, the program has been delivered and evaluated in the Mid Glamorgan probation service as the Strait Thinking on Probation Program, or STOP (Raynor and Vanstone, 1996). Probationers whose risk ratings were high were included in the study. Raynor and Vanstone (1996) compared probationers who completed the STOP program with offenders referred to other probation options and who had actuarial risk rates similar to the STOP completers. Results from the 12 month follow up indicated that the STOP completers' (n =59) actual rate of reconviction was 8 percentage points lower than their expected rates (39 percent as opposed to 47 percent respectively). The actual reconviction rates of the Other Probationers (n =100) were equivalent to their expected rates. These results, however, appear to have

diminished after two years. After two years, the STOP completers predicted rate of reconviction was 66 percent. Their actual reconviction rate was 68 percent and was closer to the actual rates of the non treatment comparison groups.

STOP completers appear to have received relatively lighter sentences upon reconviction, suggesting that they engaged in less serious crimes. Only 2 percent (1) of the STOP completers received a custodial sentence after 2 years, while 15 percent (25) of the regular probationers received a further custodial sentence. While these results are very encouraging, there was very high attrition rate within the STOP group. Thirty-eight percent of the original group dropped out of the program before completion. This could have produced a biased picture of program success.

Apart from the Welsh experience, the program has been implemented in various probation sites throughout the U.K. Data collection in almost all of these sites is restricted to consumer feedback, and in some rarer cases, the assessment of the improvement in intermediate variables that could be expected to change with participation in the program. The exception is the Swindon Probation Service where an informal recidivism outcome study was completed. James McGuire (1995b), reported on a follow up of 15 offenders involved in the Reasoning and Rehabilitation – the R & R program -- and a comparison group of 14 offenders, who were offered an intensive job search service, but were not given R & R. In the 6 month follow-up, 21 percent of the comparison group was reconvicted and 13 percent of the R & R group. More significantly, in a 12 month follow up, 64 percent (9/14) of the comparison group and 38 percent (5/13) of the R & R group reoffended. The results, however, were based on unofficial outcome data collected by

the staff at the Probation Service and the report does not provide any information on the comparability of the two groups.

For the last 5 years, the R & R program has been implemented within the British Prison Service. The Service has stressed high standards of program implementation, including the selection, training, and monitoring of the program trainers. The program has been effectively incorporated into the prison regime. To date, the Service has only preliminary outcome data from the first phase of program implementation from October 1993 to March 1996 (Cookson, 1998). The implementation of R & R coincided with the development and implementation of a Thinking Skills program, a shorter cognitive program. At the time of the evaluation, the Thinking Skills program was still under development and staff training and supervision had not reached the same high standard as it had for the R& R program. One year reconviction rates for the R&R program, the Thinking Skills program, and the Comparison group of offenders released from prison at the same time but had not received programming demonstrated a significant treatment effect. Cookson reported that the standardized reconviction rates for the 3 groups (adjusted for the effects of the covariates of previous offenses and risk score) were 21 percent for the R & R group (n=46), 38 percent for the Thinking Skills group (n=92), and 40 percent (n=857) for the comparison group. She did not, however, report on the selection criteria or the dropout rate. Although the number of subjects was low, it appears that the program was effective for all offense groups, but particularly effective for offenders over 25 years old and for the non property offenders. In Cookson's opinion, the poorer results achieved with The Thinking Skills program were because the program was not as well developed as the R&R cognitive skills program. This result, she

argued, bolstered one's confidence that the R&R results were due to treatment and not self-selection. However, there is a way in which the different program results could have been due to selection. If the R&R program produced more dropouts than the Thinking Skills program, than the results could have been completely attributable to selection. Thus, it is very important to report if attrition or self-selection is occurring, the extent of that selection, and the results of analyses which include program completers and dropouts.

B. Conclusions

On balance, there appears to be modest evidence that the Cognitive Thinking Skills program reduces criminal recidivism in general offender populations. There is stronger evidence that positive results are more likely among certain subgroups. For example, two studies suggest that the program is not effective, or not as effective, with young offenders (under 25) or with property offenders. The largest of the Canadian studies and the evaluation of the SDOP program in Colorado point to poorer results for the highest risk offenders and offenders with psychopathic characteristics, respectively. Apart from consideration of the program effectiveness based on recidivism data, several studies tapping consumers' reports (testimonials from participants and from the staff at the sites where the program has been delivered) are almost uniformly very positive. This is not a negligible consideration for administrators who are looking for initiatives that can contribute to rehabilitation goals by creating a more positive and collaborative correctional environment.

Most of the successes with the program are in probation settings. The one major study of incarcerated adults in the federal correctional system in Canada produced no treatment effect on the incarcerated group until secondary analyses were conducted. These analyses imply a further study should be done to rigorously assess the degree to which cognitive skills training may work for some adult incarcerated populations and not others. Although all studies that assessed the outcome of offenders on relevant intermediate variables reported significant improvements, even the best designed of the studies reviewed here could not satisfy Platt's requirement of a "strong inference" design that demonstrates that the positive change on pre- and post-measures is linked to reductions in recidivism. Producing significant reductions in criminal recidivism, however, is a particularly rigorous test of an intervention. Criminality is a complex, multidimensional behavior that, among high risk groups, is persistent and very resistant to change. Many mental health interventions are considered to have been effective if they simply produce improvements on intermediate variables. As Loesel observed, "meta-analyses in other fields frequently are based on "softer" outcome criteria, shorter follow-up intervals, less disturbed clients for so-called analog groups, and …treatment motivation, the milieu, and the treatment setting are at times more favorable (Loesel, 1995b)."

Based on the data, to date, which admittedly are not as strong as one would like, we can probably conclude that cognitive thinking skills program outcomes could be improved in several ways. Program effectiveness could be enhanced by screening out the most resistant offenders. Since high risk offenders pose a particular problem, a second approach is to measure and treat their needs by applying some of the recent approaches from the substance abuse field on the readiness model of

change (Miller and Rolnic, 1991; Prochaska and DiClementi, 1993), or, by providing more intensive treatment and longer term follow-up. The Cognitive Thinking Program, as designed, is not an expensive one to deliver. Non-professional staff may be trained within months. If expert monitoring and support are provided, the program can be maintained at a high standard. In Correctional Service Canada, costs per offender for completion of the program are estimated at $1,100 Canadian ($660 US). With costs this low, the treatment effects need not be large to justify the benefits of implementation of the program in correctional sites. Furthermore, the cognitive skills program could be considered a preparation or foundation course that can be used in conjunction with other intervention approaches. The interaction and cumulative effect of different offender treatment interventions may be the next major research challenge. As Lipsey's data has shown for juveniles, multimodal treatment can be quite successful. However, to date, the research on multimodal correctional treatment has not been systematic.

IV. Effect of Intensive Drug Treatment in Prison on Post-Release Outcomes

There is compelling evidence that among a significant number of individuals who commit crime, their level of crime is dramatically amplified by their use of drugs (Anglin and Speckart, 1988; Nurco, Kinlock, and Hanlon, 1990). For this reason, correctional systems have come under increasing pressure in some jurisdictions to introduce treatments that reduce or eliminate drug relapse as a way to moderate criminality once offenders have been released back to their community. Wexler, Lipton, and their colleagues have presented evidence which they claim demonstrates that therapeutic communities (TC's) or intensive residential substance abuse

treatment within prison should be the primary intervention for drug dependent offenders. In fact, Wexler has summarized the evidence on what he believes is a movement toward the use of intensive prison-based drug treatment to reduce recidivism. Among the recommendations of a panel of experts on corrections, social service, and substance abuse at a meeting sponsored by the Center for Substance Abuse Treatment was the call for the implementation of a therapeutic community "in every federal prison and every state prison system (Wexler, 1994:358)." Therapeutic communities partially isolate the drug dependent offender from the rest of the inmate population. This increases the group pressure to commit to the program and decreases the peer pressure outside of the group to maintain the lifestyle associated with their drug use.

We review the major studies which have evaluated residential, in-prison treatment. On balance, we believe the results indicate that residential treatment can moderate drug use when offenders are released to the community. The best studies in this genre support that conclusion (Martin, Butzin, and Inciardi, 1995; Pelissier, 1997; Wexler, DeLeon, Thomas, Kressel, and Peters, 1997). The effect of the intervention is strengthened when prison treatment is combined with community treatment during post-release supervision. Our enthusiasm is tempered by some of the problems we have found in this research. We have found misinterpretation of statistical analyses. More than one study has compared the outcomes of inmates treated in prison who were under post-release community supervision against untreated inmates who were under supervision for a shorter period. Since community supervision, in this context, is usually associated with lower rearrest rates, such comparisons would bias the results in favor of finding a treatment effect. We have also found studies which have compared the outcomes of treatment clients having shorter post-release

risk periods than untreated clients. This would also bias the results in favor of finding a treatment effect. These problems have received detailed treatment in Gaes (1997, 1998). However, one problem that seems to permeate this literature and many of the other substantive areas of treatment, is the biased selection of inmates. This usually occurs in one of two ways. There are procedures that select only certain inmates into an intervention program or there are processes that affect who remains in a program. In either case, researchers have compared a subset of clients who have had the ability and motivation to complete a program against comparison groups partially composed of clients who also would have completed a program and clients who would have dropped out of a program.

This approach distorts the picture we have of in-prison residential treatment outcomes. There is some evidence that in some studies, selection produces a pool of higher risk clients in treatment, while in other studies, lower risk cases were selected into treatment. Treatment dropouts also have characteristics usually associated with a higher likelihood of criminal recidivism than the average subject in a comparison group. The effect of these selection pressures is to distort the lens through which we analyze treatment effects. Some of the researchers in this area have acknowledged the selection problem and attempted to provide solutions. Because of the many pressures to place inmates in appropriate prisons, the pressure from outside influences (e.g. judges) to place particular inmates in a program, and the constant movement of the inmate population, program providers and researchers often have an extremely difficult time maintaining the integrity of an experimental or quasi-experimental design.

We are also not convinced that therapeutic communities are the only way to treat drug abusing or dependent incarcerated offenders. There is some evidence coming from evaluation of nonresidential drug treatment in Canada, that other approaches may also reduce drug relapse and recidivism, if they are carried out in conjunction with community treatment (Millson and Robinson, 1992; Millson, Weeks, and Lightfoot, 1995). The Canadian experience may be somewhat unique. CSC has a very systematic appraisal of inmate needs. This is matched by a very large menu of inmate programs intended to address many different deficits. Since drug abusers often have many deficits other than their dependency, a less intensive drug treatment regimen may work in CSC because it occurs in combination with other interventions. At this point, the CSC data are rather preliminary and it remains to be seen whether less intensive drug treatments can have any success.

Since residential treatment is the most intensive drug abuse treatment being offered in adult correctional settings, we concentrate our review on that small set of studies. Our analysis focuses on six studies that evaluated in-prison intensive drug treatment. These programs include: Stay 'N Out, Cornerstone, Key/Crest, New Vision at Kyle Unit, the Amity Right Turn Project, and a multi-site evaluation of residential drug treatment in 19 federal facilities. These programs are listed in Table 2 along with sample sizes, method of controlling for selection bias, drug and recidivism results, and comments. One of the difficulties in comparing these programs is that the specifics of the program content is rarely specified in the research studies. Other than the fact that all of these programs are residential, the actual program content is not described in any detail. Most modern residential drug treatment programs emphasize the following program components:

knowledge of drug abuse, wellness and fitness, cognitive/behavioral treatments, relapse prevention, lots of practice and role playing, techniques to increase motivation, small group sessions, and individual treatment when it is necessary. Most programs last from 6 to 12 months. There are some who also advocate the use of ex-addicts as models and trainers (Wexler, 1994). Even where program training manuals might be available upon request, it is unusual to find any measurement of the program components, or any assessment of the extent to which treatment participants internalized what they have learned. As we have advocated in other sections of this paper, such measurement would increase our understanding of successful intervention strategies and bolster our confidence in the results of quasi-experimental designs. In the following subsections, we review each program. We cite the program results and methodological problems; however, we also concentrate our critique on the procedures each evaluator used to control for selection effects.

A. Cornerstone Program

To control for selection effects, Field's (1985) evaluation of the Cornerstone Program attempted to retrospectively adjust for differences in treated and untreated clients by measuring and controlling for their criminal history. Field used indices of criminal history in an attempt to measure the similarity of individuals in different treatment groups that spent varying amounts of time in the program before they dropped out. Although Field's intent was to show that treatment duration was the key to treatment success, the analysis showed that treatment clients were very similar with respect to their criminal history background regardless of when or if they dropped out

of treatment. One would suspect that there are many different theoretical reasons a client drops out of treatment that are unrelated to the level of his or her criminal history. From our perspective, if one attempts to control for selection effects by measuring background variables, the researcher should theoretically explain why these background factors predict why clients remain or drop out of treatment.

Field's results, reported in Table 2, indicate favorable treatment outcomes. After 3 years, treatment completers had a 46 percent recidivism rate and dropouts an 85 percent recidivism rate. However, Field's evaluation was also confounded by the fact that he compared program graduates with comparison groups composed of dropouts and a group that was chosen retrospectively. The dropout rate was extremely high. In a second study, Field reported an analysis in which he noted that of 220 program admissions over a 2-year period, only 43 inmates graduated (1992: 148). With such high attrition rates and without following the program dropouts, it is extremely difficult to draw any conclusions from such a treatment evaluation regardless of how well one assesses other factors that predict recidivism that may control for some group differences.

B. New Vision Program

Initial results of the New Vision residential treatment program indicated that inmates receiving residential and aftercare treatment were less likely to recidivate (7 percent) within 6 months of release than inmates who did not receive treatment (16 percent). A design similar to Field's was

used to evaluate the program (Simpson, Knight, Chatham, Camacho, and Cloud, 1994; Knight, Simpson, Chatham, Camacho, and Cloud 1995; and Knight, Simpson, Chatham, and Camacho, in press). The program's evaluation compared a control group to a treatment group composed of inmates who participated in a 9-month prison-based TC, followed by 3 months of community-based residential treatment, followed by 1 year of outpatient treatment.

Inmates were selected into treatment through the use of a screening mechanism given to all inmates who enter Texas Department of Corrections facilities. A treatment referral committee reviewed the inmate's records, which include self-reported drug use. Inmates who had less than 9 months remaining on their sentences or who had committed an aggravated offense were excluded from further referral. Inmates who qualified for treatment had their cases forwarded to the Texas Parole Board for the final decision on placement in a drug program. Both comparison and treatment subjects in this study completed the initial referral process. The Parole Board rejected a certain number of inmates for treatment while still granting parole to these inmates. The reasons for these decisions were not specified by the authors. Thus, we have an initial selection process that differentiates treatment and comparison subjects. As it turned out, based on a composite risk assessment, treatment subjects were at higher risk for recidivism than were comparison subjects. Nevertheless, Parole Board members used their "clinical judgment" to further refine the selection process based on some unknown set of "clinical" criteria.

The attrition process for this evaluation was described comprehensively and provides a good indication of how difficult it is to conduct follow up interviews for this population once they are

released to the community. Of 482 treatment referrals, 386 (80%) graduated; 29 inmates (6%) were transferred for medical reasons, outstanding warrants, or inappropriate classification of drug problems; and 67 (14%) were terminated for program non-compliance. Unfortunately, no attempt was made to follow up on the program terminations. Also, there was attrition among those who completed the program and those who constituted the control group. At the time of the 6-month follow up, only 222 of the original 386 treatment graduates and 75 of 121 control group inmates released to parole could be interviewed. Attrition was due to offenders who moved out of the area accessible to interviewers, who were recommitted to prison, who could not be located, or who refused to be interviewed. While the outcomes of this evaluation are encouraging, the differences may be completely attributable to the selection processes operating in the program.

C. Stay 'N Out Program and Amity Right Turn Programs

Another technique that has been used to control for selection bias is to offer a program and choose volunteers who have sufficient time remaining on their sentence to complete the program. Control groups are chosen from inmates who do not have enough time to complete the program. It is unlikely that there is a relationship between the time remaining on an inmate's sentence, the level of treatment motivation, and the risk of relapse. Under these conditions, the evaluation results will generalize only to treatment volunteers. Nevertheless, the results should be unbiased if there are no other factors influencing program participation and program outcomes.

This technique has been used to evaluate the Stay 'N Out Program in New York State (Wexler

and Chin, 1981; Wexler, Falkin, and Lipton, 1988; Wexler, Falkin, and Lipton, 1990; Wexler, Falkin, Lipton, and Rosenblum, 1992; Wexler and Williams, 1986) and the Amity Right Turn Project funded by the California Department of Corrections (Wexler, DeLeon, Thomas, Kressel, and Peters, 1997). A multivariate analysis of the data from the Stay 'N Out program showed that treatment volunteers to a therapeutic community were more likely to recidivate in a post-release period than comparison volunteers receiving no treatment or less intensive treatment. Although some of the univariate results may have been more favorable, we believe the multivariate analyses were more appropriate.

The evaluation of the Amity Right Turn Project used a more complicated design (Wexler, DeLeon, Thomas, Kressel, and Peters, 1997). Inmates who had volunteered to be treated, who had a drug problem, and who were within 9 to 14 months of their parole release composed a waiting list of eligible participants. From this pool, inmates were randomly selected to participate in the prison TC. Inmates who were eligible but could not be treated prior to their release composed the control group (n=290). There were four treatment groups consisting of inmates who had been randomly selected for treatment. The composition of the four study groups depended upon whether they volunteered for post-release community-based treatment and whether they completed the prison or community-based program. Thus, the first study group was composed of inmates who volunteered for the prison program but who were terminated (prison treatment dropouts, n=95). The second study group consisted of inmates who completed the prison drug program, but did not volunteer for the community-based program (prison treatment completions, n=193). The third study group was composed of inmates who volunteered and

completed prison drug treatment and who volunteered and were terminated from the community-based program (prison treatment completions/community-based dropouts, n=45). The fourth study group was composed of inmates who volunteered and completed both the prison and the community-based programs (prison completions/community-based completions, n=92).

Wexler et. al., reported that the no-treatment control group had significantly higher reincarceration proportions at both 12 and 24 months after release from prison than all of the other study groups combined. The 12-month comparison indicated that 49.7 percent of the control group recidivated while the combined recidivism percentage of all four study groups was 33.9 percent. At 24 months these percentages were 59 and 42.6 respectively. When the combined result was separated into the control group and four study groups, the five groups had the following reincarceration percentages at 12 months: control group, 49.7; prison treatment dropouts, 45; prison treatment completions, 40; prison treatment completions/community-based dropouts, 40; prison treatment completions/community-based completions, 6.5.[1] A logistic regression of background factors in conjunction with the treatment effect indicated that reincarceration was 42 percent less likely for the combined treatment groups than for the control group.

Wexler et al. acknowledge that their results were confounded by the fact that during the post-

[1] In their report, Wexler et al. did not provide the actual percentages of inmates who were reincarcerated for the prison dropouts, prison completions, and prison completions/community-based dropouts groups. We had to estimate these percentages from a bar chart represented as Figure 1.

release period, inmates who were receiving treatment in the community-based TC were at much lower risk than other study group releasees simply by their residence in the TC. This would also affect the 24 month outcomes as well. If the risk periods were defined as beginning the day after release from the community-based facility or the day after release from prison for clients who did not participate in the community-based facility, the "risk environment" would have been more comparable for the different groups involved in the evaluation. It is clear from the analysis of the individual study groups that the dramatic differences between the combined study group and the control group was primarily attributable to the prison treatment completion/community-based completion group, the same group whose risk environment was much more benign. Although no analysis was presented, there were much more modest differences between the control group and the other three study groups composed of inmates who spent little or no time in the community-based aftercare facility.

Secondly, as the authors acknowledged, while they were able to control for selection bias during the prison treatment phase, they were unable to control for selection bias during the community-based treatment phase. Thus, although inmates were randomly assigned to the prison TC, there was a selection process operating during the community phase of treatment.

D. Key-Crest Program

The Key-Crest Program is a drug treatment intervention occurring in three phases (Martin, Inciardi, and Saum, 1995; Martin, Butzin, and Inciardi, 1995; Inciardi, Martin, Butzin, Hooper,

and Harrison, in press). The Key component is a prison therapeutic community (TC) for inmates in the Delaware corrections system. Crest, the second component, involves inmates released to a community work-release center where they maintain jobs in the community but live in a facility where they continue their drug treatment in a modified TC. In the final component, offenders are released to the community, either under parole or some other form of supervision.

The results of the evaluation, to date, suggest that residential treatment in combination with aftercare can significantly reduce drug relapse and recidivism (see Table 2). These authors controlled for selection bias by randomly assigning inmate volunteers to treatment and control groups without justifying the decision. Thus, unlike the previous study which only excluded inmate volunteers if they did not have a sufficient amount of time remaining on their sentence to complete the program, no justification was given to volunteers who were randomly assigned to a no treatment comparison group in the Key/Crest study. The Key/Crest program used such a technique for two of their four study groups. As noted in Deleon, Inciardi, and Martin (1995), the group receiving drug treatment in a work-release center was composed of some clients who did not volunteer for treatment but were motivated by the possibility of an early release. Some of these inmates "displayed negative attitudes toward the treatment program, which generally led to their quitting or being discharged from the Crest program" (1995:88). One can only suspect that inmates who volunteered for treatment and were assigned to a control group, viewed this as a capricious decision.

The research design for this program consisted of four groups. The Key group was composed of

inmates selected by correctional counselors and who volunteered to participate in the prison-based TC. Because the Crest program had not yet been implemented, these inmates were the only Key program participants who did not subsequently participate in the Crest stage. The second group consisted of Key-Crest inmates who participated in both stages. Virtually all Key graduates were allowed to participate in Crest after it was implemented. The third and fourth groups were composed of inmates who had drug abuse problems, had not participated in Key, and were given the opportunity to participate in the Crest work-release program. On a random basis, half of these volunteers were provided the Crest program (Crest-only group) while half participated in work-release in the absence of residential drug treatment (comparison group). Thus, the comparison group for these analyses were inmates who had drug abuse problems, had volunteered for Crest, and had not received in-prison TC drug treatment but had received AIDS/HIV prevention education.

An 18-month follow up of the program showed that 77 percent of Key-Crest participants reported being arrest-free, while 57 percent of Crest-only, 43 percent of Key-only, and 46 percent of the comparison group reported being arrest-free. While these results are promising, there were two selection bias processes operating in the Key-Crest program. The first selection process involved selection into the Key and Key-Crest groups. It appears that the selection depended upon staff's evaluation of candidates for the program. The second selection process occurred as a result of the way baseline data were gathered. These data were gathered just prior to an inmate's release from prison. Baseline data were collected on Key graduates, but not on Key terminations. Thus, only Key graduates were followed in the longitudinal design**.** Data were gathered on Crest and

comparison subjects at baseline in the absence of any knowledge about potential future attrition in these two groups. Thus, both the Key and Key-Crest groups were composed of inmates who were motivated enough to graduate from the Key component of this program. Even though Key-Crest participants had the opportunity in this design to drop out of the program while they were in the Crest stage, this group already was composed of a very select group of motivated individuals. However, all inmates in the Crest-only group were still followed even if they dropped out of the program (Inciardi, 1997). Despite these limitations, the data offer further support that residential treatment, in combination with aftercare, are important determinants of drug treatment success.

E. Residential Treatment in the Bureau of Prisons

Another approach that can be used to evaluate inmate programs in the presence of selection bias is to model the selection process and use this information in a simultaneous test of the selection process and the program outcome. This was the approach adopted by Pelissier (1997) to evaluate the in-prison drug treatment programs administered in 19 federal prisons. The federal residential drug treatment program consists of all of the components we listed at the beginning of this section; however, the program did not use prior inmates who had been drug abusers as trainers or treatment providers. Ninety-six percent of women and ninety-two percent of men who completed residential treatment in prison also received community aftercare. Thus, there was no explicit test of the two components of the program. Pelissier (1997) selected comparison subjects from the 19 sites where treatment was available as well as sites where there was no treatment available.

Inmates with a moderate to severe drug abuse problem could volunteer to participate in an intensive in-prison treatment program.

To analyze comparison subjects from treatment sites, Pelissier used statistical procedures to model the selection process in treatment sites while simultaneously testing the effect of treatment on post-release relapse and recidivism. These procedures have been suggested by Heckman (1979). The analysis demonstrated that there was a selection process, that this process actually would have biased the results *against* finding an effect of treatment, and that a treatment effect did occur. Multivariate survival analysis indicated that within the first 6 months of release into the community, 20 percent of the treated and 36 percent of the untreated inmates had at least one drug relapse (positive urinanalysis test). Furthermore, 3.1 percent of the treated and 15 percent of the untreated inmates were arrested on a new charge. The study also showed that one of the most important determinants of post-release criminal recidivism was whether the offender had a drug relapse. This study suggests that in-prison treatment combined with aftercare, at least in the short term may have dramatic effects. An analysis of long term findings is forthcoming.

F. Conclusions

In summary, there seems to be some evidence that intensive drug treatment in prison may reduce drug relapse and criminal recidivism in the first 6 to 18 months after release. In this area, the Key-Crest, Amity Right Turn, and Bureau of Prisons studies have the strongest research and analysis designs. In addition to the approaches used in these three studies to handle selection bias, future

studies should also measure and report two fundamental characteristics of their treatment and comparison groups -- the level of drug dependence and the level of risk of criminal recidivism. Since treatment effects depend on these characteristics, future reviews of drug treatment would benefit from an assessment of the extent to which studies are evaluating interventions with similar or different client pools. Ideally, a common measurement device for drug abuse/dependence and risk of criminal recidivism would be the best solution. However, even categorical or qualitative measures indicating whether treatment was working for low or high drug dependent clients and low or high risk of recidivism prisoners would give us a better understanding of the effectiveness of drug treatment than we currently have. Clearly, the two most compelling determinants of the effectiveness of drug treatment are the nature/intensity of the intervention and the dependence/risk of the client. Future work that characterizes the components of treatment will also allow us to more systematically compare different treatment approaches.

V. Prison Education and Work Programs

Education and work programs are the cornerstones of correctional intervention. Ironically, given the historical prominence of school and work as prison programming, the empirical data on the impact of these activities on offender's lives is limited. The evidence that is available, however, is encouraging. Despite methodological shortcomings and challenges that characterize this research, the evidence suggests that carefully designed and administered education and work programs can improve inmates' institutional behavior, reduce recidivism, and promote involvement in prosocial activities after release. This section of the paper examines the body of research that contributes to

that conclusion.

A. Prison Education and Prisoner Behavior

Labor and education programs are the oldest and most enduring of all correctional intervention methods. Education programs for inmates have been a fixture of American correctional efforts for more than 150 years. The rationale for providing education services to inmates flows directly from the educational deficits that offenders bring with them to the prison. These educational deficiencies are part of a set of failures and negative experiences that include dysfunctional family relationships, inability to obtain and hold legitimate jobs, pathological relationships with members of the opposite sex, substance abuse histories, and many other criminologically relevant markers (criminogenic needs).

Improving inmates' educational skills may reduce recidivism through several mechanisms. First, providing inmates with sufficient reading and writing skills to enable functional literacy may increase the possibility of lawful employment after release from prison. Since post-release employment is an important factor in enabling ex-offenders to remain crime free, education programming in prisons may reduce recidivism by improving job opportunities. Second, the education process *per se* may be helpful in reducing recidivism by facilitating the maturation, conscientiousness, and dedication that educational achievement requires. In this view, education may equip offenders to evaluate their environment and their decisions more thoughtfully, and therefore make better decisions that will assist them in remaining out of prison when released. A

related argument is that through exposure to the worlds of literature, mathematics, science, humanities, and the arts by way of educational programming, the offender may develop a broader frame of reference within which to evaluate life choices. The educational setting within the prison also represents an opportunity for inmates to interact with civilian employees in the context of a nonauthoritarian, goal-directed relationship.

Education programs aimed at developing basic literacy and communication skills have been offered in prisons since the earliest U.S. reformatories were established in the 1870's. These education programs were often an adjunct to the primary reform program, religious instruction. Academic and vocational education programs began to play a central role in rehabilitation programming in the 1930's, and by the 1960's, post-secondary education programs were a feature of most U.S. prison systems (Gerber and Fritsch, 1994).

Despite decades of experience with education programs in U.S. prisons, the research literature on the efficacy of education programming in reducing criminal recidivism is not well developed. In part, this is due to the fact that for decades no prison programs were formally evaluated. The development of correctional policy in the U.S. during most of the 20th century has been guided by intuition, benevolent intentions, and experience rather than by empirical analyses of what works in reducing re-offending. One of the most striking aspects of the famous research survey reported by Lipton, Martinson, and Wilks (1975) was the small number of studies available on nearly every form of correctional treatment. This lack of a solid research base concerning the efficacy of educational programs is the result of many factors, but the design and delivery of educational programs in prisons has commonly violated many of the principles of effective correctional

treatment described earlier. That is, education programs in prison have not been directed to specific criminogenic needs of offenders, have not been part of a multimodal intervention strategy, have not considered responsivity effects, have not been tailored to address the needs of offenders in different risk classifications, and have not been adequately funded to permit the high doses of educational intervention that many offenders require. For Lipton et al., informed public policy requires solid "evidence concerning the differential effects of education programs on post-release behavior (1975: 363)."

Evaluating prison education programs brings the researcher face-to-face with the thorny problems of selection bias described above in relation to other treatment programs. State legislation or correctional agency policy often dictates assignment to education programs, so random assignment to education programs is impossible. The differential motivation *for* treatment, and differential attrition *from* treatment that one observes in the context of prison education programming makes it virtually impossible to unambiguously attribute post-release recidivism to the effect of educational programming in prison. Researchers have responded to this challenge by utilizing a variety of quasi-experimental designs to assay the effects of educational programming, but these designs often fall short of the kind of empirical evidence that would support rational policy making and program development.

Gerber and Fritsch (1994) conducted a comprehensive assessment of the research literature on correctional education. They divided the studies into three subject areas: academic education (further divided into adult basic/secondary education and college education), vocational education, and social education (often called life skills training). They used three elemental criteria

to include studies in their review: the presence of a control or comparison group, use of some form of control (e.g., matching of experimental and control subjects, random assignment to program enrollment, or *post hoc* statistical control), and reporting of tests of statistical significance. Gerber and Fritsch assigned a "methodology score" to the studies they reviewed. One point each was assigned for each occurrence of the criteria discussed above.

Fourteen studies of *pre-college education* programs examined post-release recidivism as the criterion variable. Measures of program involvement varied across studies, ranging from earning the General Equivalency Diploma or completion of a prescribed academic program, to "participation" or "enrollment in academic course work" as the measure of program exposure. For example, Anderson, Anderson, and Schumaker (1988) measured educational program exposure by completion of the GED or high school diploma, or higher. Nine of the fourteen studies found educational program participation to be related to reduced recidivism. Of the seven studies that received the highest methodology score, three found no relationship between educational programming and recidivism, and four showed inverse correlations, the more education, the lower the recidivism.

Porporino and Robinson (1992) monitored 1,736 adult basic education (ABE) participants released from Canadian prisons in 1988. Among those who completed the ABE program (equivalent to completion of 8th grade), 30.1 percent were readmitted to prison during the follow up period. Recidivism was 35.5 percent among those who were released from prison before the ABE program could be completed, and 41.6 percent among those who withdrew from the ABE program. Poporino and Robinson also reported that the effects of ABE program participation was

especially effective among higher risk offenders.

In addition to recidivism measures, Gerber and Fritsch examined four studies that investigated the relationship of educational program participation and post-release employment, and two studies that examined post-release participation in education as criterion variables. Three of the four studies of post-release employment found that inmates who participated in or completed prison education programming were more likely to be employed after release. Both of the studies that examined post-release participation in education showed that inmates who participated in educational programming while confined were more likely to continue that participation in the community after release.

Gerber and Fritsch examined 14 studies of the effect of *college* programs in prisons. Again, measurement of program participation varied across studies, from simple measures of "participation," to completion of 12 college credit hours, to completion of a college degree. Overall, they found that "most studies [10 of 14] report an inverse relationship between college education and recidivism" (1994: 6). As participation in college programs increased, recidivism rates decreased. Many of the researchers who carried out these studies recognized, however, that confounding effects were substantial. For example, Thorpe, MacDonald, and Bala (1984) reported on a study of New York inmates in which earning a college degree was associated with substantially lower return to prison rates, but the investigators acknowledged that graduates may succeed because of unmeasured attributes such as "motivation" and "competence." As with the studies of basic and secondary education reviewed by Gerber and Fritsch, analyses of college programming found that participants were more likely to be employed after release (three of three

studies), were more likely to participate in additional educational opportunities after release, and that college program participants may have more favorable prison disciplinary records than non-participants.

Vocational education programming in prisons takes numerous forms, from masonry trades to computer training. Again, the premise of these programs is that the *acquisition* of vocational skills involves commitment, goal setting and motivation, and learning of technical and nontechnical knowledge, and that the acquisition of vocational skills increases ex-offenders' legitimate employment opportunities after release. Therefore, positive effects on in-prison and post-prison rule conformity is presumed. Gerber and Fritsch examined 13 studies of vocational education programs and found an inverse relationship between participation and recidivism in nine studies. Thus, participation in vocational education programs was associated with reduced recidivism rates. As an example, Saylor and Gaes (1992) investigated vocational-technical training in the Federal Bureau of Prisons and found that "inmates who received vocational training while in prison showed better 'institutional adjustment' (fewer rule violations) than those who did not receive such training, were more likely to complete stays in a halfway house, were less likely to have their paroles revoked, and were more likely to be employed" after release (in Gerber and Fritsch,1994: 8).

Gerber and Fritsch also reported on a small group of studies that probed the effect of "life skills" or *social education training*. These programs are, in many respects, even more difficult to evaluate than traditional academic or vocational education programs. First, the content of these programs varies widely. Some focus on skills needed for daily living, such as hygiene, social

interaction norms, and basic financial management. Others focus on skills such as conflict avoidance and verbal communication skills. Second, the measurement of improvement in “coping skills," “problem-solving” or “moral development” is itself subject to measurement error. Despite these important problems related to reliability and validity, a few studies claim to have documented improvement in these psychosocial dimensions. The relationship of this "personal growth" and reduced recidivism is not documented.

Taken together, Gerber and Fritsch summarized more than 90 studies of various aspects of prison education and behavioral outcomes during and after incarceration. Acknowledging that “without adequate control techniques, it is difficult to speak definitively about the impact of correctional education programs,” they concluded that “research shows a fair amount of support for the hypotheses that adult academic and vocational correctional education programs lead to fewer disciplinary violations during incarceration, reductions in recidivism, increases in employment opportunities, and to increases in participation in education upon release (1994: 11)."

A multifacited study of recidivism among more than 14,000 inmates released from Texas prisons in 1991 and 1992 underscored their conclusion (Adams, Bennett, Flanagan, Marquart, Cuvelier, Fritsch, Gerber, Longmire and Burton, 1994). The study investigated several behavioral outcomes associated with educational programming. Among other findings, this study demonstrated the importance of operationalization of the program "participation" variable. For example, when participation in Texas prison education programs was operationalized as a simple dichotomy, without reference to time spent in the program or measurable educational achievement, there was no relationship with recidivism, defined as reincarceration. When program participation was

measured by hours of program participation, however, both vocational and academic education programs yielded reduced recidivism among inmates whose exposure to the programs was greatest. For example, "inmates with fewer than 100 hours in academic programs (at the time of release) had a reincarceration rate of 25 percent compared to 16.6 percent for inmates with more than 300 hours in academic programs and 23.6 percent for inmates who did not participate in academic programs. Similarly, inmates with fewer than 100 hours in vocational programs had a recidivism rate of 22.8 percent, inmates with more than 300 hours in vocational programs had a rate of 18.3 percent, and inmates who did not participate in vocational programs had a rate of 22.4 percent" (1994: 442).

Adams et al., also found an important interaction of program exposure and offenders' needs for educational programming. Confirming the risk principle, the greatest reduction in recidivism was evidenced among inmates whose initial educational achievement levels were low and who received the highest level of exposure to educational programming. "When these two factors [were] combined, the data suggest that the recidivism rate can be reduced by about one-third if extensive services are targeted at inmates at the lowest level of educational achievement" (1994: 447). The researchers concluded by echoing the view of many correctional investigators in the two decades since publication of Lipton, Martinson, and Wilks' *(1975) The Effectiveness of Correctional Treatment*: "that correctional intervention works best when programs are matched with offenders' needs and are delivered in a concerted, purposeful manner. This point implies that correctional program administrators must be more successful in assigning inmates to programs so as to maximize the use of resources and minimize the prospect of recidivism (1994: 448-449)."

B. Prison Labor and Prisoner Behavior

Like the findings of research on prison education programs, research on prison labor is also encouraging. It appears that prison work experience operates through several mechanisms to produce better behaved prisoners during confinement, lower recidivism rates after release, and higher rates of involvement in constructive employment after release.

If education is the oldest prison treatment program, work is the oldest activity within prisons. Indeed, the much chronicled debate between the designers of the Auburn congregate system and the Pennsylvania solitary system of prison organization was essentially a conflict about the most efficient system to organize production under conditions of maximum security. And, just as offenders present deficient educational records upon entry to prison, their work histories also reflect sketchy or nonexistent employment records, few marketable skills, and an inadequate work ethic. Thus, the purpose of prison labor has always been multifaceted, and includes inculcation of positive work attitudes and the development of personal self-discipline and marketable skills. In addition to these offender-focused goals, prison work programs have sought to be economically self-sufficient (if not profitable), and to keep inmates occupied in productive activities that reduce the management dangers associated with inmate idleness. The administration of prison labor programs, and the question of whether such programs assist in reducing recidivism is complicated by the multiple goals and objectives that are sought through prison labor (Flanagan, 1989). As a research issue, prison labor also suffers from definitional ambiguity; the definition of "prison work assignment" may range from innocuous and unimportant institutional maintenance assignments to 40 hour weeks in industrial shops that approximate real world work practices. Moreover, as

prison populations have burgeoned during the last 10 years, correctional agencies have not kept pace in providing industry-like jobs for inmates. Thus, in examining the effect of prison employment on inmate behavior, the researcher must be very careful to specify the parameters of the intervention.

The paucity of empirical evaluations of the effect of prison work is indicated by the fact that Lipton, Martinson, and Wilks did not consider the area of institutional employment at all. As recently as 1984, a Committee of the Association of the Bar of the City of New York observed that "there is no empirical study showing lower recidivism rates among inmates who have participated in meaningful vocational training and/or prison industry programs (1984: 299)." In most studies, to date, the approach has been to compare recidivism rates of prisoners released after having worked in prison industry with rates for a comparison group of non-employed releasees. *Ex post facto* control of potential selection bias is usually attempted by disaggregating the work and non-work groups on recidivism-relevant variables. This is the approach that a Task Force on Correctional Industries of the Utah Department of Corrections took in finding that the 1-year return-to-prison rate for all inmates released in 1983 was 29 percent, compared to 13 percent for correctional industry participants released during the same period (State of Utah, 1984). To test for selection bias, the researchers compared the industry and non-industry groups on race, marital status, number of dependents, religion, previous occupation and work record, crime seriousness, prior arrests and incarcerations, and measures of drug and alcohol use. No significant differences were reported on these variables.

Basinger's (1985) research on offenders employed in Ohio Prison Industries employed a similar

design, but reported modest and non-significant differences in recidivism between employed and non-employed releasees. Similarly, Johnson (1984) found no significant differences between prison industry employed prisoners and those who were not employed in prison in a 2-year community follow up.

Flanagan, Thornberry, Maguire, and McGarrell (1988) studied the effect of prison industry in a large sample of New York offenders. Industry participants had worked in prison shops for at least 6 continuous months, and a comparison group was selected from inmates confined in the same facilities during the same time period. However, prison industry participants were older, were serving longer sentences, had served more time in prison, were more likely to have been employed prior to arrest, were less likely to be pre-prison drug users, and were more likely to be black or Hispanic than were members of the comparison group. Several measures of recidivism were examined. In all but one comparison, there were no significant differences between employed and non-employed inmates. When controls were imposed for recidivism-relevant differences between the groups, the recidivism rates were virtually identical. In terms of in-prison behavior, however, participation in prison industry was consistently associated with lower rates of disciplinary problems. In evaluating the finding of no significant effect of industry participation on recidivism, one must keep in mind that "employment" among the treatment group in this study lasted an average of 18 months, and was characterized by a typical work week of 26 hours.

The most comprehensive and rigorous study of the effect of prison work and vocational training was conducted by the Office of Research and Evaluation of the U.S. Federal Bureau of Prisons (Saylor and Gaes, 1992). Unlike previous studies described above, the Post-Release Employment

Project (PREP) used a prospective, longitudinal design that featured careful matching of study group participants (those who were employed and/or received vocational training while imprisoned) and control cases, and followed almost 7,000 subjects through the release experience to 12 months after discharge from prison. PREP found that the study group inmates were more likely to be employed after release, more likely to successfully complete a halfway house stay, and less likely to have their parole supervision revoked. For example, at 12 months after release, 6.6 percent of study group participants had been revoked, compared to 10.1 percent of comparison group subjects. A study of prison employment in New South Wales (McHutchison, 1995) also found that coupling in-prison work experience and work-release transitional programming produced lower recidivism rates.

Saylor and Gaes (1997) also conducted a long-term followup of the PREP study and comparison subjects. The Bureau of Prisons automated records were searched to determine which inmates had been returned to federal custody either for a new offense or a technical violation of their supervision. In 1995, when these records were searched, if an offender had not been returned to prison, he or she would have been released a minimum of 8 years without a federal reconviction. Many of these former PREP study participants had been released for 12 years without a reconviction. A failure time model was used to analyze the possible differences between study group and comparison group releasees. Although there were no significant program effects among women, there were significant differences among the men. Male inmates who had worked in prison industries were 24 percent less likely to recidivate, while those who had participated in either vocational or apprenticeship training were 33 percent less likely to recidivate throughout the followup period.

Evaluating the value of prison-based education and work programs in achieving correctional policy goals is a complicated, tenuous undertaking. The prison experience is made up of many diverse elements. In relation to other features of the prison environment, education and work experience may be a small part of the offender's institutional career. The studies reviewed here indicate, however, that public investment in prison work and education programs may be a wise investment, and considering the total cost of recidivism, a cost-effective investment. When considered as a body of developing scientific work on the impact of prison programs, one must conclude that education and work programs can contribute significantly to increasing offenders prospects for success. Moreover, the research, to date, provides correctional authorities with a set of empirically derived guidelines for the design and delivery of such interventions. From a public policy perspective, the evidence suggest that a retreat from the public commitment to investment in prison labor and effective education programs in prison would be misguided.

VI. Treatment of Sex Offenders

Many correctional agencies are being challenged to offer more programming and improved services to sex offenders. Many of these programs are designed to be consistent with the principles outlined at the beginning of this paper. Consistent with the risk principle, Gordon, Holden, and Leis (1991) have argued that sex offender treatment should be concentrated on those who require it the most. Such a recommendation is fraught with ethical and legal questions, not the least of which is trying to decide which category of victim is harmed the most and which type of sex offender is the most at risk to re-offend.

Of the four intervention subdomains we have examined closely, sex offender treatment is the most difficult to summarize. While there is some data to suggest treatment may reduce recidivism, there are also many qualifications to this conclusion. The sex offender population is heterogenenous, yet many researchers fail to consistently classify and distinguish subgroups of sex offenders. This makes it difficult to know whether treatment failures or successes are comparable across studies. We describe these classification, definition, and treatment nomenclature problems in the first part of this section. We then review the methodological problems and describe the results of a meta-analysis of the sex offender literature, and discuss some promising studies.

A. Problems in Sex Offender Classification, Mechanisms of Change, and Treatment Nomenclature

The sex offender literature consists of a diverse collection of studies on exhibitionists, rapists, child molesters, and incest offenders. There is evidence that sex offenders are more likely to "specialize" in certain kinds of deviant sexual behavior than other offenders, who are classified on the basis of their conviction offense. The extent to which this heterogeneous group of sexual offense “subtypes” overlap in treatment studies is difficult to determine. Researchers must begin to be more systematic in their classification of sex offender subtypes. Some researchers have been very specific in categorizing client subtype and treatment modality (Hagan, King, and Patros, 1994; Knight and Prentky, 1990; Lang, Pugh, and Langevin, 1988; Marshall and Barbaree, 1988), suggesting different interventions may be more appropriate for different client subtypes. But, in general, these studies often fail to distinguish offenders by their conviction offense. When researchers do classify sex offenders, they often use different criteria. For example, study participants have been classified not only on the basis of type of conviction but also on the basis of

sexual preference or measures of deviant sexual arousal. The absence of uniformity in operational definitions makes comparison of research studies imprecise, since it is not at all clear that different treatment studies are examining the same or even a similar sex offender population.

Another source of confusion in this literature arises out of the mechanisms of change proposed in various studies. These include such diverse mechanisms as: denial, also called minimization and rationalization (Barbaree, 1991), attitudes which are supportive of deviant sexual activity, cognitive distortions (Murphy, 1990), social competence skills (Stermac and Quinsey, 1986), sexual identity, offender victimization, victim awareness, deviant arousal and fantasy (Laws and Marshall, 1990; Quinsey and Earls, 1990), anger management/impulse control (Prentky and Knight, 1986), and relapse prevention (Pithers, 1990). Which of these mechanisms are relevant for which types of sex offenders? Which mechanisms are the most important in sex offending? Are there combinations of mechanisms that act in concert that are highly correlated to sexual offending? These are only some of the questions that researchers will have to address before we can make sense of this treatment area.

It is also important to specify in which setting sex offender treatment is conducted (e.g., residential, outpatient) at what levels of intensity (e.g., duration, focus), employing which treatment techniques (e.g., cognitive-behavioral, pharmacological, psychotherapeutic) and using which modalities (e.g., individual, group). Consequently, any coherent review of the sex offender treatment literature would have to organize studies according to these characteristics. Since many studies fail to report even the most fundamental aspects of treatment, the studies are difficult to review systematically.

B. Problems in Methodology

Within a corrections environment, the random assignment of sex offenders to either "treated" or "non-treated" groups is especially problematic. While some sex offenders who are not motivated to receive treatment do not willingly participate, many service providers question the ethics of denying programming to sex offenders who wish to participate but cannot because of the research design (Marshall, 1996). Furthermore, many studies use inappropriate control or comparison groups or lack fundamental matching procedures (Baxter, Motiuk, and Fortin, 1995). For example, studies may compare treatment dropouts to treatment completers, or inmates who deny their sex offending problems with those who acknowledge them, or incarcerated sex offenders to those on probation.

Studies of sex offender treatment effectiveness should match treated offenders with untreated offenders on a set of relevant characteristics such as release date, age at release, and sentence length (Motiuk, Smiley, and Blanchette, 1996). Ideally, the control or comparison group would also be matched with the treated group on risk factors such as history of sexual offending and victim age/gender preferences. These factors have been found to be related to re-offending among sex offenders (Hanson and Bussiere, 1996). This presents yet another methodological hurdle to overcome as selection criteria for treatment could have adverse impacts on the ability to conduct matching procedures. This problem is identical to the one we described in detail in our assessment of drug treatment programs.

Different outcome measures and variable post-treatment follow up periods are typical in this area

of treatment research. Post-release outcome studies rarely concur on recidivism rates, in part because of varying definitions of what constitutes 'recidivism' (Freeman-Longo and Knopp, 1992). Treatment outcome measures have included self-reports of new offenses, charges, convictions, or returns to custody. More stringent definitions of recidivism consider only new convictions for sex crimes as an outcome measure. Again, the absence of uniformity in measures makes comparisons difficult.

Because of low base rates of sexual reoffending (Hanson and Bussiere, 1996), sample sizes need to be exceedingly large. Moreover, reliance on officially recorded convictions underestimate actual sexual recidivism rates because a great deal of sexual offending goes unreported by vulnerable or ashamed victims. (Weinrott and Saylor, 1991). The problem is further compounded by sample attrition since individuals are removed from the study or follow up for a variety of reasons (Blanchette, 1996). Other methodological problems include detailing the therapeutic intervention under investigation, measurement of the service provider's adherence to the treatment protocol, and factoring in the delay between treatment completion and release.

C. Sex Offender Treatment Results

Consistent with the problems that permeate this literature, Quinsey, Harris, Rice, and Lalumiere, (1993) have questioned whether it is possible to draw any conclusions about the effectiveness of sex offender treatment to reduce sexual reoffending over extended time periods. However, others have argued that some treatments can be empirically demonstrated to be effective with sex offenders and are, in fact, successful in reducing sexual reoffending (Barbaree, Seto, and Maric,

1996; Marshall, 1996; Robinson, 1996). Marques, Day, Nelson, and West (1994) have reported on an extremely promising evaluation of an experimental sex offender treatment program in California that appears to be methodologically sound. Unfortunately, no conclusive results are available yet.

Hall (1995) produced a meta-analysis of recent sex offender treatment studies which showed a small, but robust, effect size for sex offender treatment. Hall (1995) found that recidivism for untreated sex offenders was 27 percent compared with 19 percent for treated sex offenders, a relative reduction of 30 percent. Gordon and Nicholaichuk (1996) reported a 24-percent decrease in recidivism for sex offenders receiving cognitive-behavioral treatment. In Robinson's (1996) study of the cognitive skills training program that we have already summarized, he reported a 58-percent reduction in the recidivism of sex offenders who completed treatment while in prison. Although sex offenders appeared to achieve the greatest treatment gains from cognitive skills training relative to other offense groups (violent, drug, property), about one-third had received sex offender treatment before participating in cognitive skills training. While this is suggestive, because of the problems we have noted with this study, future research will have to disentangle whether multiple treatments can enhance the success of intervention and whether there is a particular sequence of treatments that must be followed.

While there is some research that suggests there may be a modest effect of sex offender treatment in prison, we should be very wary of drawing any sweeping conclusions. There is certainly no definitive approach to treatment. Across the many jurisdictions in Canada and the United States, sex offenders are required to complete a variety of different programs before being considered for

release. Then, they may be required to participate in maintenance programs upon their release to the community. As yet, the full impact or relative contribution of post-program efforts (i.e., relapse prevention) to reducing recidivism among sex offenders remains largely untested (Miner, Marques, Day, and Nelson, 1990).

VII. Future Directions for Correctional Research

There have been major theoretical and methodological advances in the juvenile and adult correctional treatment literature since Lipton, Martinson, and Wilk's assessment nearly a quarter century ago. We have the rudiments of a theory of intervention which treats criminal behavior as an extension of normative or prosocial behavior. Within the context of psychological learning models, several researchers have proposed a set of principles to be used to extend the social learning model to criminal behavior. Some of these principles need further clarification and empirical assessment. We also recommend more extensive use of strong inference designs. In this section, we discuss some of the theoretical shortcomings. Then we present a model of strong inference designs that we hope will clarify the importance of measuring mediating processes in an area as complicated as correctional treatment.

A. Theoretical Shortcomings

1. Problems with the Risk Principle

The risk principle, as far as we can tell, has two corollaries. The first is an empirical proposition:

"...the effects of treatment typically are found to be greater among higher risk cases than among lower risk cases. (Andrews et. al., 1990: 374)." The second proposition is prescriptive "...and involves the idea of "matching levels of treatment services to the risk level of the offender. (Andrews and Bonta, 1994:175)." As we have shown in several substantive sections of this paper, there are counterexamples to the risk principle as an empirical proposition. There are instances where lower risk offenders benefit more from a treatment than higher risk offenders. The principle also has been represented as an interaction between the level of risk and the level of treatment. High intensity interventions work better with high risk offenders, while low intensity interventions work better with low risk clients. To assess this interaction more effectively, more work must be done to classify the intensity of treatments, since there are already several risk scales to predict the likelihood of recidivism.

2. Needs Principle

The needs principle is very intuitive. There are specific human deficits directly related to the propensity to commit crime and these should be addressed. Beyond this level of intuition, however, is the need to specify and classify these needs. Researchers must demonstrate in intervention studies that these human deficits are being treated and that treatment of these deficits leads to reductions in recidivism.

3. Responsivity

Andrews and Bonta (1994) argue that the responsivity principle needs much more development. If

this principle is to move beyond the level of tautology, namely that responsivity is demonstrated by effective treatments, then a great deal of work on the classification of client learning styles and treatment provider techniques has to be done.

While theory development in correctional treatment is critical to its advancement, we believe that greater attention to strong inference designs will be just as important.

B. Strong Inference Designs

Too few studies use what Platt (1964) called a "strong inference" design. A strong inference design requires the scientist to articulate clearly both the theoretical reasons and under what conditions a behavioral change (outcome) is expected to occur. The scientist should propose alternative hypotheses or mechanisms of change. A crucial experiment or test should then be conducted. From our perspective, a crucial test includes measurement of the theoretical mechanism that hypothetically causes a change in outcome, development of procedures that ensure the scientific reliability and validity of the measurement of these intervening causal mechanisms, and, finally, proof that changes in these mechanisms are associated with changes in outcomes. Even when a design employs random assignment of clients to conditions, strong inference is bolstered by measurement of the mechanism that mediates change. When the design is quasi-experimental, measuring the mechanism of change is even more important. The principles of correctional treatment help us delineate mechanisms of change. The measurement of risk, treatment intensity, specific needs, and specific treatment styles and techniques will allow us to draw stronger inferences about the mechanisms of behavior change and the effectiveness of

treatment.

Figure 1 represents some of the possible inferences that can be deduced when one uses a strong inference design of a correctional treatment evaluation. Three related questions can be asked. Is there a positive change in the pre- to post-measure of the theoretical mechanism that the social scientist proposes is the source of change in the outcome? Is the magnitude of change in the pre- to post-measure of the theoretical mechanism correlated to the level of change in outcome? Finally, is the level of outcome for the treatment group better than the level of outcome for the control or comparison group? In Figure 1, we have sketched out some, but not all of the possible paths that lead to different levels of inference depending on the answers to these questions. For example, the strongest inference comes from an affirmative answer to all three of these questions. Using these questions also raises other theoretical possibilities. For example, consider the following path: there is an indication that there was a pre- to post-difference in the level of the proposed change mechanism and the magnitude of change in the pre- to post-measure of the change mechanism was positively correlated to the level of change in outcome; however, there was no significant difference in the outcome levels for treatment and control groups. This may suggest the researcher was on the right track, but that the level of treatment was not strong enough or something else weakened the impact of treatment. There are many other inferences suggested by the different paths and we have clearly not considered all of the paths or all of the possible insights suggested by the different paths. However, it is an extremely rare instance in which researchers use strong inference designs in the correctional treatment evaluation literature.

Summary

We have found support for the effectiveness of correctional interventions; however, some interventions seem to work better than others and they work better on juveniles than on adults. We have noted serious definitional, analytical, and methodological problems in each of the specific areas of adult correctional treatment we have reviewed. Are these problems serious enough to vitiate any general conclusions we might make about these subdomains of intervention? These were the same circumstances encountered by members of the NAS Panel on Rehabilitative Techniques and by Lipton, Martinson, and Wilks (1975). Although their conclusions were summarily represented as "nothing works," as many others have pointed out, a closer reading of these publications indicates that authors of these studies thought there were promising intervention programs. We are in a position to make a stronger statement, namely that correctional treatment for adults has modest but substantively meaningful effects. Even though the level of recidivism is modest, even small reductions can produce future reductions in criminality. It is probably premature to conduct cost-benefit analysis comparing the gains of treatment against the benefits and costs of alternative crime reduction policies for prisoners and probationers. Unfortunately, there is still a great deal of variability in both the quality of the programs and the quality of the evaluation research.

In more recent years, the results of a number of meta-analyses have been interpreted to suggest that despite methodological problems, there may be types of interventions that can reduce adult post-release recidivism. Even though most of these studies involve juvenile interventions and the effects are modest, some types of interventions seem to have a dramatic impact on post-release behavior. The next decade of treatment evaluations should capitalize on the results of the best meta-analyses of correctional treatment. Using the principles suggested by meta-analysis and a

coherent theory of treatment, researchers can test these principles under rigorous evaluation conditions, using strong inference designs, and insuring there are sufficient replications. This presents the greatest opportunity to advance the science of correctional treatment. We would be better positioned to properly assess the evidence, if more of the studies were better designed and executed. We also think that correctional practice could be improved if the underlying scientific foundation of correctional intervention research were improved.

We must also come to terms with the unwieldy assortment of programs we currently call treatments. We need a system to classify treatments according to the behavior the intervention is intended to modify, the mechanism of change, and those factors that mediate, inhibit, or catalyze change. Perhaps theorists can build upon the principles of correctional treatment and catalog different interventions according to these principles. The aim of this classification is to get a clearer understanding of essential mechanisms of change and those which are peripheral. For example, are literacy or cognitive skills fundamental to any other intervention? Are there certain emotional or motivational deficits that must be addressed before any program can be successful? The classification of client and treatment characteristics will also benefit future meta-analyses. It is not always clear that researchers are classifying certain treatments into the same categories.

We also need to understand the boundaries and relative impact of individual and collective interventions. Intuitively, it would seem some programs address behavioral change at the margin while others strike at the heart of the problem. This calls for research designs that include treatment manipulations and the measurement of post-release contextual factors. For example, training inmates in specific work skills may have little or no impact if they return to economically

and socially impoverished conditions. A test of inmate work training should include an analysis of economic and social conditions in different release jurisdictions. If the next decade of correctional treatment research is more imaginative, systematic, and rigorous, the next generation of meta-analyses should produce more demonstrative effects.

References

Adams, K., Bennett, K. J., Flanagan, T. J., Marquart, J. W., Cuvelier, S. J., Fritsch, E., Gerber, J., Longmire, D. R., and Burton, Jr.. V. S. 1994. "A large-scale multidimensional test of the effect of prison education programs on offenders' behavior." *The Prison Journal* 74(4): 433-449.

Anderson, S. L., Anderson, D. B., and Schumacker, R. E. 1988. *Correctional Education: A Way to Stay Out.* Springfield: Illinois Council on Vocational Education.

Andrews, D. A. 1989. "Recidivism is predictable and can be influenced: Using risk assessments to reduce recidivism." *Forum on Corrections Research*, 1(2): 11-18.

Andrews, D. A. 1995. "The psychology of criminal conduct and effective treatment," in McGuire, J. (ed.), *What Works: Reducing Reoffending*, 35-62, John Wiley and Sons: New York.

Andrews, D.A. 1996. "Recidivism is predictable and can be influenced: An update." *Forum on Corrections Research*, 8, (3): 42-44.

Andrews, D. A. and Bonta, J. 1994. *The Psychology of Criminal Conduct*, Cincinnati, Ohio: Anderson Publishing Co.

Andrews, D.A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., and Cullen, F. T. 1990., "Does correctional treatment work? A psychologically informed meta-analysis," *Criminology*, 28: 369-

404.

Anglin, M. D. and Speckart, G. 1988. "Narcotics and crime: A multi-sample, multimethod analysis." *Criminology*, 26: 197-233.

Antonowicz, D. and Ross, R. R. 1994. "Essential components of successful rehabilitation programs for offenders." *International Journal of Offender Therapy and Comparative Criminology*, 38: 97-104.

Association of the Bar of the City of New York, Committee on Corrections 1984. "Can our prisons become factories with fences?" *The Record of the Association of the Bar of the City of New York* 40(4): 299.

Barbaree, H. E. 1991. "Denial and minimization among sex offenders: Assessment and treatment outcome." *Forum on Corrections Research*, 3 (4): 30-33.

Barbaree, H.E. Seto, M.C. and Maric, A. 1996. "Effective sex offender treatment: The Warkworth Sexual Behavior Clinic." *Forum on Corrections Research*, 8 (3): 13-15

Basinger, Aaron 1985. "Are prison work programs working?: the impact of prison industry participation on recidivism rates in Ohio." MPA Thesis. Columbus: The Ohio State University.

Baxter, D.J., Motiuk, L. L., Fortin, S. 1995. "Intelligence and personality in criminal offenders."

In D. H. Saklofske and M. Zeidner (eds.), *International Handbook of Personality and Intelligence* New York: Plenum.

Begg, C.B. 1994. "Publication bias," in H. Cooper and L. Hedges (eds.), *The Handbook of Research Synthesis*, 399-410, New York: Russel Sage Foundation.

Blanchette, K. 1996. *Sex Offender Assessment, Treatment and Recidivism: A Review of the Literature*. Research Report, Ottawa: Correctional Service of Canada.

Cookson, H. 1998. "Impact of cognitive skills programmes on reconviction rates." *Living Skills Newsletter*, London: H. M. Prison Service.

Cooper, H. and Hedges, L.V. 1994. "Potentials and limitations of research synthesis." in Cooper, H. and Hedges, L. V. (eds.) *The Handbook of Research Synthesis*, New York: Russel Sage Foundation.

Cooper, H. and Hedges, L. V. 1994. *The Handbook of Research Synthesis*, New York: Russel Sage Foundation.

DeLeon, G., Inciardi, J. A., and Martin, S. S. 1995. "Residential drug treatment research: Are conventional control designs appropriate for assessing treatment effectiveness?" *Journal of Psychoactive Drugs.* 27(1): 85-91.

Dickersin, K. 1994. “Research registers.” in H. Cooper and L. V. Hedges (eds.) *The Handbook of Research Synthesis*, 71-84, New York: Russel Sage Foundation.

Elliot, D. S. 1997. Editor’s introduction. In D. S. Elliot (Ed.) *Blueprints for Violence Prevention (Vols. 1-10)*, Boulder, Colorado: Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado.

Field, G. 1985. "The Cornerstone program: A client outcome study." *Federal Probation*, 49(2): 50-55.

Field, G. 1992. "Oregon prison drug treatment programs." *Drug Abuse Treatment in Prisons and Jails*. National Institute on Drug Abuse Monograph, 118. U.S. Department of Health and Human Services, 142-155. Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.

Flanagan, T. J. 1989. "Prison labor and industry." In Lynne Goodstein and Doris L. MacKenzie (Eds), *The American Prison: Issues of Research and Policy*. New York: Plenum Publishers.

Flanagan, T. J., Thornberry, T. P., Maguire, K., and McGarrell, E. 1988. T*he Effect of Prison Industry Employment on Offender Behavior: Final Report of the Prison Industry Research Project.* Albany, NY: University at Albany, State University of New York.

Freeman-Longo, R. E. and Knopp, H. F. 1992. "State-of-the-art sex offender treatment: outcome

and issues." *Annals of Sex Research*, 5 (3): 141-160

Gaes, G. G. 1998. "Correctional treatment." in M. Tonry (ed.) *The Handbook of Crime and Punishment*, 712-738, New York: Oxford University Press.

Gaes, G. G. 1997. "A Review of Recent Studies of High Intensity Adult Correctional Drug Treatment Programs: The Problem of Selection Bias and Possible Solutions," Unpublished Manuscript, Federal Bureau of Prisons, Washington, D.C.

Garrett, C. J. 1985. "Effects of residential treatment on adjudicated delinquents: A meta-analysis." *Journal of Research in Crime and Delinquency*. 22(4): 287-308.

Gendreau, P. and Andrews, D. A. 1990. "Tertiary prevention: What the meta-analysis of the offender treatment literature tells us about 'what works'." *Canadian Journal of Criminology*, 32: 173-184.

Gendreau, P., and Ross, R. R. 1979. "Effectiveness of Correctional treatment: Bibliotherapy for cynics." *Crime and Delinquency*, 25: 463-489.

Gendreau, P., and Ross, R. R. 1987. "Revivification of rehabilitation: Evidence from the 1980's." *Justice Quarterly*, 4(3): 349-407.

Gensheimer, L. K., Mayer, J. P., Gottschalk, R., and Davidson, W. S. 1986. Diverting youth from

the juvenile justice system - A meta-analysis of intervention efficacy." In S. J. Apter and A. P. Goldstein (Eds.) *Youth Violence*. 39-57. New York: Pergamon Press.

Gerber, J. and Fritsch, E. J. 1994. "The effects of academic and vocational program participation on inmate misconduct and reincarceration." Chapter 3 in Sam Houston State University, *Prison Education Research Project: Final Report.* Huntsville, TX: Sam Houston State University.

Gordon, A., Holden, R., and Leis, T. 1991. "Managing and treating sex offenders: Matching risk and need in programming." *Forum on Corrections Research*, 3 (1): 7-11.

Gordon, A. and Nicholaichuk, T. 1996. "Applying the risk principle to sex offender treatment." *Forum on Corrections Research*, 8(2): 36-38.

Gottschalk, R., Davidson, W. S., Mayer, J., and Gensheimer, L. K. 1987a. "Behavioral approaches with juvenile offenders: A meta-analysis of long-term treatment efficacy." In E. K. Morris and C. J. Braukman (Eds.) *Behavioral Approaches to Crime and Delinquency: A Handbook of Application, Research, and Concepts*. 399-423, New York: Plenum Press.

Gottschalk, R., Davidson, W. S., Gensheimer, L. K., and Mayer, J. 1987b. "Community-based interventions." In H. C. Quay (Ed.) *Handbook of Juvenile Delinquency*. 266-289. New York: Wiley and Sons.

Gould, S. J. 1981. *The Mismeasure of Man*, New York: W. W. Norton & Company

Hagan, M.P., King, R. P., and Patros, R. L. 1994. "The efficacy of a serious sex offenders treatment program for adolescent rapists." *International Journal of Offender Therapy and Comparative Criminology*, 38 (2): 141-150.

Hall, G. C. 1995. "Sex offender recidivism revisited: A meta-analysis of recent treatment studies." *Journal of Consulting and Clinical Psychology*, 63 (5): 802-809.

Hall, J. A., Tickle-Degnen, L., Rosenthal, R., and Mosteller, F. "Hypotheses and problems in research synthesis," in H. Cooper and L. Hedges (eds.), *The Handbook of Research Synthesis*, 17-28, New York: Russel Sage Foundation.

Hanson, R. K. and Bussiere, M. T. 1996. "Sex offender risk predictors: A summary of research results." *Forum on Corrections Research*, 8 (2): 10-12.

Heckman, James 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153 - 61.

Inciardi, J. A., Martin, S. S., Butzin, C. A., Hooper, R. M., and Harrison, L. D. In press. "An effective model of prison-based treatment for drug-involved offenders." *Journal of Drug Issues.*

Inciardi 1997. Personal communication.

Izzo, R. L. and Ross, R. R. 1990. "Meta-analysis of rehabilitation programs for juvenile

delinquents." *Criminal Justice and Behavior*, 17: 134-142.

Johnston, G. and Hunter, L. 1992. "Evaluation of the specialized drug offender program." *Report for the Colorado Judicial Department*, Bolder: University of Colorado, Center for Addiction Research.

Johnston, G. and Hunter, L. 1995. "Evaluation of the specialized drug offender program." In Ross, R.R. and Ross, R. D. (eds.) *Thinking Straight: The Reasoning and Rehabilitation Program for Delinquency Prevention and Offender Rehabilitation*, Ottawa: Air Training and Publications.

Johnson, C. M. 1984. "The Effects of Prison Labor Programs on Post-Release Employment and Recidivism." Doctoral dissertation. Tallahassee: Florida State University.

Knight, K., Simpson, D. D., Chatham, L. R., Camacho, L. M., and Cloud, M. 1995. "Prison-based treatment assessment (PTA): Final report on 6-month follow up study." Fort Worth: Texas Christian University, Institute of Behavioral Research.

Knight, K., Simpson, D. D., Chatham, L. R. and Camacho, L. M. In press. "An assessment of prison-based drug treatment: Texas' in-prison therapeutic program." Fort Worth: Texas Christian University, Institute of Behavioral Research.

Knight, R. A. and Prentky, R. A. 1990. "Classifying sex offenders: The development and corroboration of taxonomic models." In W. L. Marshall and H. E. Barbaree (Eds.), *Handbook of*

Sexual Assault: Issues, Theories, and Treatment of the Offenders 23-52. New York: Plenum.

Lang, R. A., Pugh, G. M., and Langevin, R. 1988. "Treatment of incest and pedophilic offenders: A pilot study." *Behavioral Sciences and the Law*, 6: 239-255.

Laws, H. E. and Marshall, W. L. 1990. "A conditioning theory of the etiology and maintenance of deviant sexual preference and behavior." In W. L. Marshall and H. E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offenders*. 209-229. New York: Plenum.

Lipsey, M. W. 1989. "The efficacy of intervention for juvenile delinquency: Results from 400 studies," Paper presented at the 41st meeting of the American Society of Criminology, Reno, NV.

Lipsey, M. W. 1992. "Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects," in Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., and Mosteller, F. (Eds.) *Meta-Analysis for Explanation: A Casebook*, 83-128, New York: Russel Sage Foundation.

Lipsey , M. W. 1995. "What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents?" in McGuire, J. (ed.), What Works: Reducing Reoffending, 63-78, John Wiley and Sons: New York.

Lipton, D. 1998. "How do cognitive skills training programs for offenders compare with other

modalities: A meta-analytic perspective." Paper presented at the Stop and Think Conference, York, England.

Lipton, D., Martinson, R., and Wilks, J. 1975. *The Effectiveness of Correctional Treatment.* New York: Praeger.

Lipton, D. S., Pearson, F. S., Cleland, C., and Yee, D. 1998. "Synthesizing correctional treatment outcomes: Preliminary findings from CDATE" New York: National Development and Research Institutes, Inc.

Loesel, F., Koferl, P. and Weber, F. 1987. *Meta-Evaluation of Social Therapy*. Stuttgart, Germany: Ferdinand Enke Verlag.

Loesel, F., and Koferl, P. 1989. "Evaluation research on correctional treatment in West Germany: A meta-analysis." In H. Wegener, F. Loesel, and J. Haisch (Eds.) Criminal Behavior and the Justice System: Psychological Perspectives. New York: Springer.

Loesel, F. 1993. "The effectiveness of treatment in institutional and community settings." *Criminal Behavior and Mental Health*. 3: 416-437.

Loesel, F. 1995a. "The efficacy of correctional treatment: A review and synthesis of meta-evaluations," in McGuire, J. (ed.), *What Works: Reducing Reoffending*, 79-114, John Wiley and Sons: New York.

Loesel. F. 1995b. "What recent meta-evaluations tell us about the effectiveness of correctional treatment." In Davies, G. and Lloyd-Bostock, S. (eds.), *Psychology, Law, Criminal Justice: International Developments in Research and Practice*,. Berlin: Walter DeGruyter.

Loesel, F. 1996. "Effective correctional programming: What empirical research tells and what it doesn't." *Forum on Corrections Research*, 8 (3): 33-37.

Logan, C. H., Gaes, G. G., Harer, M. D., Innes, C. A., Karacki, L. and Saylor, W. G. 1991. Can meta-analysis save correctional rehabilitation. Washington, D.C.: Federal Bureau of Prisons.

Marques, J. K., Day, D. M., Nelson, C., and West, M. 1994. "Effects of cognitive-behavioral treatment on sex offender recidivism: Preliminary results of a longitudinal study." *Criminal Justice and Behavior*, 21 (1): 28-54.

Marshall, W. L. 1996. "Assessment, treatment, and theorizing about sex offenders: Developments during the past twenty years and future direction." *Criminal Justice and Behavior*, 23 (1): 162-199.

Marshall, W. L. and Barbaree, H. E. 1990. "Outcome of comprehensive cognitive-behavioral treatment programs." In William L. Marshall, D. R. Laws, and H. E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender*. 363-385. New York: Plenum Press.

Martin, S. S., Butzin, C. A., and Inciardi, J. A. 1995. "Assessment of a multistage therapeutic community for drug-involved offenders." *Journal of Psychoactive Drugs,* 27(1): 109-116

Martin, S. S., Inciardi, J. A. and Saum, C. A. 1995. "Strengths and shortcomings of prison-based treatment for male drug-involved offenders: An empirical assessment." Presentation at the Meetings of the Academy of Criminal Justice Sciences, Boston, Massachusetts.

Martinson, R. 1974. "What works? Questions and answers about prison reform," *The Public Interest* 35:22-45.

Mayer, J., Gensheimer, L. K., Davidson, W. S. II, and Gottschalk, R. 1986. "Social learning treatment within juvenile justice: A meta-analysis of impact in the natural environment." In S. J. Apter and A. P. Goldstein (Eds.) *Youth Violence: Programs and Prospects*. 24-57, New York: Pergamon Press.

McGuire, J. 1995a. *What Works? Reducing Reoffending*. New York: Wiley

McGuire, J. 1995b. "Community-based reasoning and rehabilitation programs in the UK." In Ross, R. R. and Ross, R. D. (eds.) *Thinking Straight: The Reasoning and Rehabilitation Program for Delinquency Prevention and Offender Rehabilitation*, Air Training and Publications. Ottawa, Canada

Miller, W. R. and Rollnic, S. 1991. *Motivational Interviewing: Preparing People to Change*

Addictive Behavior. New York: Guilford Press.

Millson, B. and Robinson, D. 1992. *An assessment of the offender substance abuse pre-release program at Drumheller Institution*, Ottawa, Correctional Service of Canada

Millson, W. A., Weekes, J. R., and Lightfoot, L. O. 1995. *The offender substance abuse pre-release program: An analysis of pre-release and post-release outcomes*. Ottawa, Correctional Service of Canada

Miner, M. H., Marques, J. K., Day, D. M., and Nelson, C. 1990. "Impact of relapse prevention in treating sex offenders: preliminary findings." *Annals of Sex Research*, 3 (2): 165-185.

Morris, N. 1998. Plenary Address at the National Workshop on Assessing the Effectiveness of Corrections Programs. Corrections Program Office, Office of Justice Programs, February 10-12, Chicago.

Motiuk, L. L., Smiley, C., and Blanchette, K. 1996. "Intensive programming for violent offenders: A comparative investigation." *Forum on Corrections Research*, 8 (3): 10-12

Murphy, W. D. 1990. "Assessment and modification of cognitive distortions in sex offenders." W. L. Marshall, D. R. Laws, and H. E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender*. 331-342. New York: Plenum Press.

Nuffield, J. 1982. *Parole Decision-Making in Canada: Research Towards Decision Guidelines*. Ottawa: Supply and Services.

Nurco, D. N., Kinlock, T. W., and Hanlon, T. E. 1990. "The drugs-crime connection." In J. A. Inciardi (Ed.) *Handbook of Drug Control in the United States* 71-90. Westport, CT: Greenwood Press.

Palmer, T. 1992. *The Re-emergence of Correctional Intervention*, Sage : Newbury Park Calif.

Palmer, T. 1975. "Martinson revisited." *Journal of Research in Crime and Delinquency*, 12:133-152.

Pelissier, B. 1997. Triad drug treatment evaluation project: Six-month interim report. Unpublished Manuscript, Federal Bureau of Prisons, Washington, D.C.

Pithers, W. D. 1990. "Relapse prevention for sexual aggressors: A method for maintaining therapeutic gain and enhancing external supervision." In W. L. Marshall, D. R. Laws, and H. E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender* 343-361. New York: Plenum Press.

Platt, J. R. 1964. "Strong inference," *Science*, 146, No. 3642. 16 October, 1964.

Porporino, F. J. Fabiano, E. A. and Robinson, D. 1991. "Focussing on successful reintegration:

cognitive skills training for offenders." Research Report #19. Research and Statistics Branch, Ottawa: Correctional Service of Canada.

Porporino, Frank and D. Robinson 1992. *Can Educating Adult Offenders Counteract Recidivism?* Ottawa: Correctional Service Canada.

Prentky R. A. and Knight, R. 1986. "Impulsivity in lifestyle and criminal behavior of sexual offenders." *Criminal Justice and Behavior*, 13: 141-164.

Prochaska, J. O. and DiClementi, C. C. 1986. "Toward a comprehensive model of change." In Miller, E. E. and Heather, N. (eds.) *Treating Addictive Behaviors: Processes of Change* . New York: Plenum Press.

Quinsey, V. L. and Earls, C. M. 1990. "The modification of sexual preferences." In W. L. Marshall, D. R. Laws, and H. E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender* 279-295. New York: Plenum Press.

Quinsey, V. L., Harris, G T., Rice, M. E. and Lalumiere, M. L. 1993. "Assessing treatment efficacy in outcome studies of sex offenders." *Journal of Interpersonal Violence*, 8 (4): 512-523.

Raynor, P. and Vanstone, P. 1996. "Reasoning and rehabilitation in Britain: The results of the straight thinking on probation (STOP) programma." *International Journal of Offender Therapy and Comparative Criminology* 40(4): 272-284.

Redondo, S., Garrido, V., and Sanchez-Meca, J. 1996. "Is the treatment of offenders effective in Europe?: The results of a meta-analysis." Presented at the 48th Annual Meeting of the American society of Criminology, Chicago, Il. November 20-23.

Roberts, A. R. and Camasso, M. J. 1991. "The effect of juvenile offender treatment programs on recidivism: A meta-analysis of 46 studies." *Norte Dame Journal of Law, Ethics, and Public Policy*. 5: 421-441.

Robinson, D. 1995. "The impact of cognitive skills training on post release recidivism among Canadian federal offenders." Research Report, Research Branch, Ottawa, Correctional Service Canada.

Robinson, D. 1996. "Factors influencing the effectiveness of cognitive skills training. *Forum on Corrections Research*, 8 (3): 6-9

Robinson, D. 1998. Personal Communication.

Robinson, D. Grossman, M., and Porporino, F. 1991. " Effectiveness of the cognitive skills training program: From pilot to national implementation." Research and Statistics Branch, Ottawa. Correctional Service of Canada.

Rosenthal, R. and Rubin, D. B. 1982. "A simple, general purpose display of magnitude of experimental effect." *Journal of Educational Psychology*, 74(2): 166-169.

Ross, R. R. and Fabiano, E. 1985. "Time to think: A cognitive model of delinquency prevention and offender rehabilitation." Johnson City, NN: Institute of Social Sciences and Arts.

Ross, R.R., & Fabiano, E., and Ewels, C. D. 1988. "Reasoning and rehabilitation." *International Journal of Offender Therapy and Comparative Criminology*, 32 29-35.

Saylor, William G. and Gerald G. Gaes 1992. PREP Study Links UNICOR Work Experience with Successful Post-Release Outcome. U.S. Department of Justice, Federal Bureau of Prisons *Research Forum* 1(3): 1-8.

Saylor, William G. and Gerald G. Gaes 1997. "Training inmates through industrial work participation and vocational and apprenticeship instruction." *Corrections Management Quarterly*, 1(2): 32-43.

Sechrest, L. White, S. O. And Brown, E. D. 1979. *The Rehabilitation of Criminal Offenders: Problems and Prospects*, National Academy of Sciences, Washington, D.C.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., and Bushway, S. D. 1998. *Preventing Crime: What Works, What Doesn't, What's Promising*. National Institute of Justice, Washington, D. C., URL: http//www.preventingcrime.org/

Simpson, D. D., Knight, K., Chatham, L. R., Camacho, L. M., and Cloud, M. 1994. "Prison-based treatment assessment (PTA): Interim report on Kyle IT evaluation." Presented to the Texas

Commission on Alcohol and Drug Abuse. Fort Worth: Texas Christian University, Institute of Behavioral Research.

State of Utah, Governor's Task Force on Correctional Industries 1984. *The Challenge of Correctional Industries.* Salt Lake City: Utah Department of Corrections.

Stermac, L. E. and Quinsey, V. L. 1986. "Social competence among rapists." *Behavioral Scientist*, 8: 171-285.

Thorpe, T., D. MacDonald and G. Bala 1984. "Follow-up of Offenders Who Earned College Degrees While Incarcerated in New York State." *Journal of Correctional Education* 35(3): 86-88.

Tobler, N. 1986. "Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group." *Journal of drug Issues*. 16(4):537-567.

Weinrott, M. R. and Saylor, M. 1991. "Self-report of crimes committed by sex offenders." *Journal of Interpersonal Violence*, 6 (3): 286-300.

Wexler, H. K. 1994. "Progress in prison substance abuse treatment: A five year report" *The Journal of Drug Issues*, 24(2): 349-360.

Wexler, H. K. and Chin, J. 1981. "Evaluation of a system to evaluate prison-based drug treatment." Paper presented at the Eastern Evaluation Research Society Conference, Philadelphia, Pennsylvania.

Wexler, H. K., Falkin, G. P. and Lipton, D. S. 1988. "A model prison rehabilitation program: An evaluation of the Stay'n Out Therapeutic Community." Final report to the National Institute on Drug Abuse. NCJ-108560. New York: Narcotic and Drug Research, Inc.

Wexler, H. K., Falkin, G. P. and Lipton, D. S. 1990. "Outcome evaluation of a prison therapeutic community for substance abuse treatment." *Criminal Justice and Behavior,* 17(1): 79-92.

Wexler, H. K., Falkin, G. P., Lipton, D. S., and Rosenblum, A. B. 1992. "Outcome evaluation of a prison therapeutic community for substance abuse treatment." In *Drug Abuse Treatment in Prisons and Jails*. 246-260. National Institute on Drug Abuse Monograph, 118. U.S. Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.

Wexler, H. K. and Williams, R. 1986. "The Stay'n Out therapeutic community: Prison treatment for substance abusers." *Journal of Psychoactive Drugs*, 18(3): 221-230.

Wexler, H. K., DeLeon, G., Thomas, G., Kressel, D., and Peters, J. 1997. "The Amity prison TC evaluation: Reincarceration outcomes." Unpublished manuscript, Center for Therapeutic Community Research at the National Development and Research Institutes. Inc.

Whitehead, J. T., and Lab, S. P. 1989. "A meta-analysis of juvenile correctional treatment." *Journal of Research in Crime and Delinquency*. 26:276-295.

Table 1. Meta-analyses of juvenile and adult correctional treatments.			
Study	Number of Studies Included in the Meta-analysis	Juvenile	Adult
Andrews et al., 1990	80	Yes	Yes
Garrett, 1985	111	Yes	No
Gensheimer et al., 1986	44	Yes	No
Gottshalk et al., 1987a	90	Yes	No
Gottshalk et al., 1987b	25	Yes	No
Izzo and Ross, 1990	46	Yes	No
Lipsey, 1992; 1995	443	Yes	No
CDATE -- Lipton et al., 1998	900	Yes	Yes
Loesel et al., 1987; 1989; 1993	8	No	No
Mayer et al., 1986	34	Yes	No
Redondo, 1996	57	No	Yes
Roberts, 1991	46	Yes	No
Tobler, 1986	143	Yes	No
Whitehead and Lab, 1989	50	Yes	No

Table 2. Residential Drug Treatment Programs:

Sample size (N), Nature of Control Group, Recidivism and Relapse Results, and Comments

Program	N	Nature of Control Group	Follow Up Period	Results: Criminal Recidivism	Results: Drug Relapse	Comments
Corner-stone	567	Dropouts & Offenders from another Jurisdiction	3 years	Prison Tx[2]: 46% Dropout Control: 85% Control Grp2: 64%	Not Reported	Retrospective design One treatment site
New Vision	297	Volunteer Controls	6 months	Prison Tx: 7% Control: 16%	Prison Tx: 38% Control: 55%	Comparisons based on Treatment Completers High Attrition Rates One treatment site
Stay 'N Out	1,428	Volunteer Controls	Variable Up to 3 Years	Prison Tx: 27% Control: 41%	Not Reported	Multivariate Results were ambiguous One treatment site
Amity Right Turn	715	Volunteer Controls	12 Months	Prison Tx & Aftercare: 6.5% Prison Tx Completers/ Aftercare Dropouts: 40% Prison Tx Completers: 40% Prison Tx Dropouts: 45% Intent to Treat Controls: 34% Controls: 34%	Not Reported	Appeared to be different risk periods for the different groups. One treatment site
Key-Crest	457	Random Assignment for the Aftercare Residential Phase	6 Months	Prison Tx Only: 18% Aftercare Only: 15% Prison Tx & Aftercare: 5% Control: 28%	Prison Tx Only: 30% Aftercare Only: 15% Prison Tx & Aftercare: 3% Control: 65%	Recidivism and relapse data based on interviews of inmates Prison group consisted of only treatment completers One treatment site

[2] Tx is used as an abbreviation for Treatment

Bureau of Prisons	1,800	Randomly Selected Drug Abusers in Sites Where No Treatment was Available	6 months	Prison Tx: 3% Control: 15%	Prison Tx: 20% Control: 35%	Aftercare provided to most program participants, therefore difficult to disentangle its unique contribution Evaluation included 19 different treatment sites with the same treatment regimen.

Appendix I. Limitations and Problems of the Treatment Meta-analysis Literature

This appendix examines some of the problems associated with meta-analysis in the context of the treatment literature. The discussion begins with the interpretation and coding of studies, the most fundamental part of the procedure. We also discuss the possible biased selection of studies and research registers which are a way to ensure that all studies, published and unpublished will be readily available to the scientific community.

A. The Reliability of Studies Coded for the Purpose of Meta-analysis

Andrews, Zinger, Hoge, Bonta, Gendreau, and Cullen (1990) completed a meta-analysis based upon the theoretical formulation of effective treatment outlined at the beginning of this paper and articulated in Andrews and Bonta (1994). In addition to the effect sizes, a moderator variable was coded on the basis of whether or not a study followed one of four treatment regimens: criminal sanctions -- the total absence of rehabilitative programs; inappropriate correctional service -- rehabilitative programs that were not guided by the principles outlined at the beginning of this paper; appropriate correctional service -- rehabilitation based on delivering services to higher risk cases, using behavioral programs reflecting responsivity, or non-behavioral programs that were structured and targeted criminogenic need; and finally, unspecified correctional service -- treatment that could not be categorized as appropriate or inappropriate. Andrews et al., found that the average effect size for appropriate correctional treatments implied that such treatments could reduce recidivism by 50 percent. The other types of treatment either had a weak influence (unspecified) or were associated with higher recidivism (criminal sanctions, inappropriate

treatment). While this meta-analysis was theoretically guided, it did have several problems. Unlike the Lipsey meta-analysis, few other methodological or contextual moderator variables were coded. Lipsey has shown that methods variables account for about half of the variation in effect sizes. One would assume, that had Andrews et al., been just as comprehensive in their coding, they would have found a more modest average treatment effect.

A second problem with this study points to another caution researchers must observe when they conduct meta-analyses. The treatment regimen variable coded by Andrews et. al., which specified interventions as criminal sanctions, appropriate correctional service, inappropriate correctional service, and unspecified correctional service was subject to too much variation in interpretation. In an attempt to replicate the findings of the Andrews et al., (1990) meta-analysis, Logan, Gaes, Harer, Innes, Karacki, and Saylor (1991) read and evaluated all of the studies considered by Andrews et al., to be exemplars of appropriate correctional treatment. Logan et al, found that many of these studies contained insufficient detail to allow an inference about the study quality. Some arbitrary decisions seemed to be made in selecting comparisons that demonstrated appropriate treatment. In some cases, effect sizes were based upon secondary analyses of the data conducted by the original researchers, when there had been no primary effect of treatment. These post hoc analyses are considered data snooping by most methodologists and statisticians and often result in findings that cannot be replicated. This is because, if you do enough post hoc comparisons, eventually, you will find results by chance alone. To understand how difficult it is to code these kinds of moderator variables, the researcher has to get into the minutiae of the methods section of every study. Logan et al, (1991) have shown that a separate set of eyes can come to different conclusions about the interpretation of moderator variables and the effect of

treatment. Rather than debate whether Andrews et. al., or Logan et al., were more correct in their interpretation of the outcomes of the studies, it is more important to focus on the methods researchers use in approaching the meta-analysis.

There are two primary lessons that can be learned from this. The broader the definition of a variable (as was the case with the Andrews et al., definition of appropriate and inappropriate treatment), the less reliability there will be in the coding of such variables. A second, more subtle lesson is also evident. If meta-analysis is conducted by advocates or proponents of a particular perspective, the research should have additional safeguards to protect against biased interpretation of the outcomes of a study. Norval Morris calls this "partisan research" (1998). Stephen J. Gould (1981) has demonstrated how subtle bias can be in the interpretation of even physical measurement when the researcher believes in a particular point of view. Researchers with a strong theoretical orientation should use others to assess the quality and consistency of their coding, as well as report the reliability of those who have been tasked with coding the studies.

B. Biased Selection of Studies and the Use of Research Registers

Yet another potential problem of meta-analysis is the possibility of bias in the selection of studies. Publication bias can take several forms. Reporting bias results from the failure by authors to report statistically non-significant results. There can also be retrieval bias. When the researcher collects studies to synthesize, it is much more difficult to gather unpublished studies or even discarded studies which represent statistically non-significant results. We can only speculate whether it is more or less likely that there is reporting bias in correctional treatment research than

in other research domains.

Begg (1994) has examined this issue in some detail using medical research. Citing other studies of publication bias, Begg reported evidence that the odds of having a study published was as high as 3.4 to 1, if the study demonstrated statistical significance, as opposed to non-significance. The lowest odds he found were 1.8 to 1 in favor of studies finding statistical significance for an effect. As Begg asserts, "results of the meta-analysis should be very compelling before we regard them as definitive (p.408)." One way to begin to address the problem of publication bias is to establish research registers.

A research register is a compilation of all initiated research projects in a particular domain (Dickersin, 1994). It can be a database of planned, ongoing, or completed studies. Dickersin's field is medical research. Her inquiries and analyses seem to be confined to that research domain. Nevertheless, her insights may prove to be equally valuable, if not more so, to the social sciences, and to program and treatment evaluation in particular. As Dickersin indicates, registers are a way to overcome publication bias. She argues that most research syntheses in medicine were based on published studies and estimated that only 50 to 80 percent of all studies had been published. She also cites five studies that she claims demonstrated “consistent and persuasive evidence for the existence of publication bias (p.75)”, whether the designs were observational or experimental. Although the existence of research registers may not completely solve the problem of publication bias, if we make registers a common practice in social science research in general and program evaluation research in particular, we can begin to have greater confidence in our coverage of all studies initiated, ongoing, and completed in a particular research domain. Obviously, there are a

great many practical difficulties to overcome. Nevertheless, it would seem that if certain social science domains are going to develop, a research register may be a significant step in their maturity. Members of the CDATE project spent a great deal of time and effort to collect published and unpublished studies. Gathering unpublished studies retrospectively is extremely difficult. A treatment research register might make future comprehensive meta-analyses more thorough and perhaps easier to do. Research registers can be updated to include the results of meta-analyses so that there is a common public access data source for everyone to examine the studies that have been started or completed and the calculation of effect sizes and the coding of moderator variables for a particular evaluation. Such a public domain dataset can reduce the level of publication bias and allow different researchers to compare their assessment of a given study against someone else's meta-analytic results.