Evidence-based Practices in Corrections

A Training Manual for the California Program Assessment Process (CPAP)

Office of Research
California Department of Corrections and Rehabilitation

Center for Evidence Based Corrections
University of California, Irvine

Memorandum

Date   :   February 1, 2006

To   :   Roderick Hickman, Secretary
                Jeanne Woodford, Undersecretary
                Joe McGrath, Chief Deputy Secretary, Adult Institutions
                Del Sayles-Owen, Chief Deputy Secretary, Adult Programs
                Bernard Warner, Chief Deputy Secretary, Juvenile Justice

Subject: **CALIFORNIA PROGRAM ASSESSMENT GUIDELINES AND TRAINING MANUAL**

This manual provides training materials for the proposed California Program Assessment Process (CPAP). It was assembled by the California Department of Corrections and Rehabilitation, Office of Research, and the University of California, Irvine, Center for Evidence-based Corrections. CPAP is a tool for determining whether offender change programs are likely to succeed in delivering on the CDCR's promise to reduce recidivism through evidence-based crime prevention and recidivism reduction strategies. The CPAP gives the CDCR an objective and consistent method for evaluating the design and implementation of programs using the best available evidence regarding what design elements make correctional programs effective in reducing recidivism.

Correctional agencies across the nation looking for ways to incorporate evidence-based methods into their correctional practice have sought guidance from the National Institute of Justice (NIJ) and a commercially available program assessment instrument called the Correctional Program Assessment Inventory (CPAI). Both NIJ and the CPAI are in general agreement about program principles that are most likely to prevent recidivism. The CPAP is designed around these same principles, which are as follows:

1. Programs should have a clearly articulated and research-based model.
2. Programs should target offenders who are the greatest risk to re-offend.
3. Programs should address criminogenic needs.
4. Programs should be delivered at the highest reasonable level of intensity.
5. Programs should be responsive to the temperament, learning style, motivation and culture of offenders.
6. Programs should use positive incentives and appropriate sanctions.
7. Programs should produce continuities between program activities and pro-social offender support networks.
8. Program staff training and education should be appropriate and adequate for the purpose of the program.
9. Programs should measure performance and use that information for continuous improvement.

In addition to evaluating these critical programmatic elements in both existing and proposed correctional programs, the CPAP provides a method of assessing the strength of existing research around particular programs. Programs that have been shown to reduce recidivism through rigorous research designs will be given higher priority than those without such evaluations.

The CPAP uses an instrument designed by the CDCR Office of Research, in conjunction with the Center for Evidence-Based Corrections, to evaluate proposed and existing CDCR programs that aim to reduce recidivism. Implementing CPAP assessment of proposed and existing recidivism-reduction programs will allow the CDCR to:

- Maintain objectivity and consistency in the approving proposed programs and prioritizing and replicating existing programs.
- Minimize the influence of opinion, ideology, and lobbying on which programs receive CDCR support.
- Communicate clear and coherent standards for what constitutes a well-designed and evidence based correctional program.
- Guide program directors and staff in improving the design and implementation of programs.

Ryken Grattet, Ph.D.
Assistant Secretary (A), Office of Research

Jesse Jannetta, MPP
UCI Center for Evidence-based Corrections

Jeff Lin, Ph.D.
UCI Center for Evidence-based Corrections

Table of Contents

INTRODUCTION

In its strategic plan, published in January of 2005, the California Department of Corrections and Rehabilitation (still known at that time as the Youth and Adult Correctional Agency) announced a new mission statement:

To improve public safety through evidence-based crime prevention and recidivism reduction strategies

This short statement presents two ideas with profound and far-reaching implications for CDCR's program activities.
1. The work of CDCR is aimed at reducing the recidivism of its charges, and the success of the agency must be judged against that standard.
2. The measures the CDCR brings to bear in pursuit of this goal must be based on evidence, not on opinion, ideology, or untested views.

CDCR defines "evidence-based practice" in the field of corrections as:

> The conscientious, explicit, and judicious use by correctional administrators of current best research evidence in selecting programs designed to manage offenders, reduce recidivism, and increase public safety. Research evidence of program effectiveness must adhere to accepted methodological standards. A program must also use empirical assessment tools to target the program to the individual offender and must objectively measure program implementation and outcomes. Evidence-based programs also includes programs that adhere to 'principles of effective intervention' established by prior research.

In order to ensure that the CDCR's programming for offenders is consistent with evidence-based practice to the greatest extent possible, the CDCR employs the California Correctional Program Assessment Process (CPAP) to assess rehabilitative programs that aim to reducing offender recidivism. This training manual is designed to prepare CDCR personnel to perform the role of rating proposed or existing programs on their conformity to evidence-based practice, using the CPAP rating scale, devised by the CDCR Office of Research and the University of California, Irvine's Center for Evidence-Based Corrections. This manual describes the CPAP and explores in detail the CPAP rating instrument, as well as the meaning behind the rating categories, issues that may occur in evaluating programs, and the proper application of the scoring rules.

Thank you for agreeing to serve as a CPAP rater and for contributing to the safety of California citizens through enhancing the effectiveness of the CDCR's work.

The CPAP Rating Instrument

The CPAP rating instrument has three components, the Effective Interventions Scale, the Research Basis Scale, and the Program Fidelity Scale. Proposed programs are rated using the first two scales. All three scales are applied to rate programs that have been implemented. The CPAP is designed to be applicable to programs for youth and adults, housed in CDCR facilities or under CDCR supervision in the community.

The Effective Interventions Sale

The items rated under the Effective Interventions Scale are based on CDCR's Principles of Effective Change Programs (see Box 1), which are derived from academic research into corrections, and reports prepared by the National Institute of Corrections and the National Institute of Justice.

Box 1: Principles of Effective Change Programs

> 1. **Programs should have a clearly articulated and evidence-based model.**
> 2. **Programs should target offenders who are the greatest risk to re-offend.**
> 3. **Programs should address criminogenic needs.**
> 4. **Programs should be delivered to high risk offenders at the highest reasonable level of intensity.**
> 5. **Programs should be responsive to the temperament, learning style, motivation and culture of offenders.**
> 6. **Programs should use positive incentives.**
> 7. **Programs should produce continuities between program activities and pro-social offender support networks.**
> 8. **Program staff training and education should be appropriate and adequate for the purpose of the program.**
> 9. **Programs should measure performance and use that information for continuous improvement.**

There is a consensus among researchers and corrections professionals that each of the items contained in this scale is important for an effective recidivism-reduction program. The definition and scoring rule for each item is discussed in detail in Sections 4 through 8.

Effective Interventions Scale (EIS)

Item	Scoring Rule	Pts
1. Risk Assessment at pre-program phase	A. Program conducts or relies on a risk assessment instrument to determine supervision requirements and to determine the appropriateness of the program to the risk level of the offender. To receive points a program must conduct "meaningful" risk assessment. That is, there must be consequences resulting from the use of the tool. Fulfilling this requirement means that offenders can be excluded from the program based upon the assessment. Programs that target "high risk" offenders receive more credit than those that	3

		target medium and low risk offenders (See "risk principle"). A program must use an assessment tool that has been shown reliable and valid in previous research. A program that relies on risk assessment conducted elsewhere counts (i.e., the program need not have its own risk assessment as long as it relies on the classification applied by another program or the CDCR). If there is no logical reason why a risk assessment should be used then this category should be omitted from the total score and labeled "NA" for not applicable. (2 pts for use of risk assessment, 1 pt for targeting high risk offenders)	1
2.	Needs Assessment at pre-program phase	A. Program conducts or relies on a needs assessment instrument to determine services required by the offender. To receive points a program must conduct "meaningful" needs assessment. That is, needs identified by the assessment must align with a case plan. A program must target the criminogenic needs of the offender. A program must use an assessment tool that has been shown reliable and valid in previous research. A program that relies on needs assessment conducted elsewhere counts (All or nothing)	4
3.	Program Model	A. The program must be based on a clearly articulated theoretical model that links the intervention content directly to an offender's criminogenic need	2
		B. Program manual or curriculum materials exist (all pts or none)	2
		C. Uses cognitive behavior or social learning methods (all pts or none)	2
		D. Program enhances intrinsic motivation	1
		E. Program is structured to produce continuities between the program activities and communities, families, and other programs (1 pt for coordination with communities, 1 pt for either coordination with families or other programs)	2
		F. Program dosage varies with offender risk level (Higher risk offenders receive greater program dosage)	1
		G. Program design reflects the responsivity principle (i.e., it has procedures to determine the preparedness of the offender for the program and to match the delivery of the program to the learning style of the offender).	1
		H. Program design identifies positive reinforcement strategies, not just sanctions	1
4.	Program Administration	A. 75% or more of service staff possess an undergraduate degree. Among those with degrees, 75% of staff has degrees in a helping profession (1 pt. for each)	2
		B. 75% of staff have worked in offender treatment programs for at least two years	1
		C. Explicit strategy for recruitment and retention of staff	1

	D. Initial training on program model, including written materials	1
	E. Program director was involved in the design of the program, has at least 3 years of experience with offenders, and has a degree in social work or related field (1 pt ea.)	3
5. Quality Assurance	A. Program collects data to monitor performance (1 pt), includes individual level data on participation (1 pt), identifies the eligible population (1 pt), data is forwarded and analyzed by a non-program entity (1 pt)	4

Research Basis Scale

This scale allows the raters to score programs on the persuasiveness of the research evidence supporting a given program model. The meaning and scoring of the specific items on this scale are discussed in detail in Sections 9 and 10 of this manual. The scale itself is structured as follows.

Research Basis	
Item	*Pts*
1. An expert committee, respected advisory group, or Best Practices panel recommends	+1
2. Multiple positive evaluations exist	+2
3. Negative or no effect evaluations	-1
4. Published in peer reviewed outlet	+2
5. Research Rigor	
A. Level 1: Correlation between program participation and recidivism reduction, temporal sequence between program participation and recidivism reduction clearly observed, or a comparison group present without demonstrated comparability to the treatment group (& no controls)	+1
B. Level 2: A comparison between two or more units of analysis, one with and one without the program (with partial controls)	+4
C. Level 3: A comparison between multiple units with and without the program, controlling for other factors, or a nonequivalent comparison group has only minor differences evident	+6
D. Level 4: Random assignment and analysis of comparable units to program and comparison groups	+10

(Note: "+10" appears spanning items B–D at the right of the Research Rigor block.)

Program Fidelity Scale

A program designed using evidence-based principles must be implemented in accordance with its design. To measure the alignment between program design and implementation, the Program Fidelity Scale rates the extent to which existing programs are implemented in conformity to the principles of effective intervention. This scale is used in tandem with the Effective Interventions Scale only after a program has been implemented. Points are deducted from a program's score on the Effective Interventions Scale, because the value of fidelity to the model depends upon the quality of that model. So, for example, a program might receive 4 points for its use of needs assessment, but 0 points for its use of risk assessment at the proposal evaluation stage. In that case, the program could only have 4 points deducted during the program implementation evaluation stage for shortcomings on its implementation of assessment.

As each of the items on this scale is tied to one of the items in the Effective Interventions Scale, the meaning and scoring of each item of this scale will be discussed in Sections 4 through 8 below.

Program Fidelity Scale (Subtracted from Effective Intervention points)		
Item	*Scoring Rule*	*Pts*
Assessment Use	Greater than 10% departures from assessment tools. This relates to how determinative the assessment tools are for placement in the program.	-4
	Program combines high and low risk offenders. The "risk principle" is based on research that shows that individuals of different risk levels should not be combined within treatment programs and that interventions should be directed at high risk offenders, since low risk offenders are less likely to re-offend	-3
Program Model	Implementation not aligned with the program model	-6
Program Structure	Continuities between in-prison, reentry, and community supports not well effectuated (e.g., below 50% of the offenders are linked across domains)	-2
	Program dosage not linked to risk level	-1
	Responsivity assessment not determinative of service/program delivery	-1
	Positive reinforcement absent	-1
Program Administration	Staff training/educational requirements relaxed in practice	-2
	Staff turnover high (greater than 25% per year)	-3
Quality Assurance	Inconsistent reporting	-2
	Eligibility requirements not adhered to	-1

The Rating Process and Procedures

Eligible Programs

Programs for CDCR institution and parole populations can be divided into three types[1]:

1. Change—Programs intended to change the antisocial behavior of offenders/wards through rehabilitation, education, and employment programs. Rehabilitative programs are designed to address the immediate sources of an offender's criminal behavior and include substance abuse and psychological treatment programs. Educational and employment programs are designed to impart tools that can be used by the offender/ward to avert further criminal behavior.
2. Activity—Programs that occupy the time of the ward/offender, stimulate them, or engage them in pro-social activity to ease their adjustment to custody or to the community after release. Some examples include yoga, reading groups, support groups, programs designed to assist parolees with finding health care or reuniting with family members.
3. Control—Programs designed to increase offender/ward compliance and enhance the supervision capabilities of correctional officials. Global Positioning System (GPS) monitoring and electronic in-home detention are good examples.

Although all three programs types are valuable and necessary, change programs are the most mission-centered and the CDCR will not be able to deliver on its promise of improved public safety without effective change programs. Therefore, change programs must be evaluated by means of the CPAP, so that CDCR resources can be reserved for the programs with the greatest likelihood of effectiveness in reducing recidivism.

The CPAP Review Process

Program providers who wish to propose a change program to the CDCR, for either the custody or parole population, may submit a proposal according to the submission guidelines in Box 2. The CPAP rating group will then rate the program on the Effective Interventions and Research Basis Scales. The CPAP rating group consists of representatives from the Office of Policy Analysis and Planning, the Office of Research, the Division of Adult Programs, and the Division of Juvenile Programs. The CPAP rating group will convene as needed to review proposed programs. The group will assess existing programs based upon the request of the Secretary, Undersecretary, or any of the Chief Deputies.

The CPAP rating group works on the consensus model to arrive at agreement on the scores to be assigned for each item. There is no "passing" threshold score; rather the CPAP rating scores are designed to allow CDCR to allocate program resources according to the relative strengths of existing and proposed programs, and also to allow program proponents to gauge their program design against the CDCR's principles for evidence-based practice. CDCR's Office of Research, the Office of Policy Analysis and Planning,

---

[1] It is possible for a single program to combine elements from more than one program type.

and the juvenile and adult program staff will work with program proponents to assist them in developing and revising proposed or existing programs to better conform to the CPAP principles. The CPAP rating system is meant to be a general instrument, and there may be programs that do not fit neatly into the rating categories of the scoring instrument. In that event, the program can receive an "NA" ("not applicable"). Scores on each scale will be reported a percentage basis to adjust for items that are "NA" for certain types of programs.

**Box 2: Submission Guidelines for Program Proposals
California Department of Corrections and Rehabilitation**

Program proposals to the CDCR should include an executive summary or a project abstract, and address each of the following questions:

Program Model
1. What are the goals of the program?
2. What is the treatment approach employed by the program to meet the goals?
3. What is the research evidence supporting the program's treatment approach?

Program Eligibility and Assessment
1. How is program eligibility determined?
2. What is the risk level (high, medium, low) of the eligible offender population? Will a standardized instrument be used to assess risk? (If so, please attach.)
3. What criminogenic needs does the program seek to address? Will a standardized instrument be used to assess needs? (If so, please attach.) How will the program use that assessment information?

Program Structure
1. What are the program requirements for participants?
2. How much time does an offender devote to the program? (Length of sessions, number of sessions per week, duration of the program in months) Does this vary by offender risk level?
3. What methods do program staffers utilize to support and encourage offender change? (Behavior modeling, motivational interviewing, social learning, etc.)
4. How does the program respond to individual differences in offender learning style, level of motivation, level of maturity, cultural background, and other relevant differences in receptiveness?
5. How does your program enhance offender motivation?
6. What is the program reward and sanctions structure?
7. How does the program create continuities between program activities and offenders' pro-social support networks?
8. Please submit your program manual and/or curriculum material.

Staff Qualifications/Selection/Training
1. How are program staff/volunteers selected and trained?
2. Please submit résumés for the program director and other relevant staff.
3. Please submit samples of training material to be used for staff and/or volunteers.

Measurement and Evaluation
1. What data will you collect regarding offender assessment and case management?
2. How will you measure incremental offender change while they are participating in the program? (e.g., ongoing risk and needs assessments, data tracking, etc.)
3. What are your program outcome measures, and how will you track them?
4. How will you measure staff performance?
5. How will you use the data that you collect for program improvement?
6. Are you planning an outside evaluation of program effectiveness?

**EFFECTIVE INTERVENTIONS AND
PROGRAM FIDELITY SCALES**

1. Risk Assessment at Pre-Program Phase

"A program conducts or relies on a risk assessment instrument to determine supervision requirements and to determine the appropriateness of the program to the risk level of the offender. To receive points a program must conduct "meaningful" risk assessment. That is, there must be consequences resulting from the use of the tool. Fulfilling this requirement means that offenders can be excluded from the program based upon the assessment. Programs that target "high risk" offenders receive more credit than those that target medium and low risk offenders. A program must use an assessment tool that has been shown reliable and valid in previous research. A program that relies on risk assessment conducted elsewhere counts. If there is no logical reason why a risk assessment should be used then this category should be omitted from the total score and labeled "NA" for not applicable."

EIS 1A. "The risk principle" states: "Programs should target offenders who are the greatest risk to re-offend." By targeting the highest risk offenders for change programs, CDCR can allocate its resources to working with the offenders who represent the greatest risk to the public and who are likely to require the most intensive program intervention if recidivism is to be prevented.

In order to conform to the risk principle, three conditions must hold:
1. A risk assessment instrument must be used.
2. The risk assessment instrument must be valid.
3. The risk assessment must be meaningful.

A risk assessment instrument must be used to determine the risk level of potential program participants. "Risk" in this context means risk to commit criminal offenses. A risk assessment instrument evaluates the offender for factors that have a demonstrated relationship to likelihood of future offending. The risk measured is an actuarial risk, meaning that it predicts the probability that an offender with certain characteristics will re-offend. Such instruments are not designed to predict with certainty whether specific individuals will re-offend.

Relevant risk factors may be both "static" (unchanging, such as commitment offense, history of substance abuse, and mental illness), or "dynamic," (or subject to change, such as antisocial attitudes and job skills.) To the extent that dynamic factors are important predictors of risk, it is desirable to assess offender risk regularly, to capture the changing impact of dynamic risk factors on an offenders overall risk level. It is not necessary that the proposed program conduct its own risk assessment. It may rely on risk assessments done elsewhere.

Risk assessment instruments are valid if they have been demonstrated to effectively predict risk of recidivism, generally through a process of formal evaluation. Ideally, the risk assessment instrument will not only have been validated, but have been validated specifically for the CDCR offender population, or any subset of that population targeted

by a program. For example, there are several risk assessment instruments designed specifically for sex offenders. The predictive accuracy of specific risk factors may vary across geographic areas or due to other differences between populations, so the very best risk assessment instruments are those that have been tested on the specific program population, and failing that, one as similar to the program population as possible.

Conducting a risk assessment of a program population has little value if the results of that assessment are not applied in the operation of the program. A risk assessment is said to be meaningful if it has reliable consequences for program placement. The program must make clear what offender level of risk will be included in the program, and must exclude offenders that are not assessed at that level of risk. Programs should also avoid mixing risk levels within the program. Combining high and low risk offenders has been linked to higher recidivism for the low risk offenders. Offenders of different risk levels can participate in the same program, but only if the program has separate tracks or sessions for the different risk levels.

> **Key Questions**
> - What risk assessment instrument is being used on the program population? How will the program ensure that all participants have been assessed for risk level?
> - Does this program target the highest-risk offenders? If not, is there a valid reason not to do so?
> - Are offenders with different levels of risk kept separate by the program?

Scoring Rule

A program receives may receive 3 points for use of risk assessment. 2 points are awarded for use of a valid risk assessment, and 1 point for targeting high risk offenders. A risk instrument is considered valid if there has been any validation study of it within a five year period using California data. Tests of either predictive or construct validity are acceptable. A longer period can be accepted if the case can be made that the inmate or parole population has not fundamentally changed since the last validation. For programs serving female offenders, only use of risk assessment instruments that have been validated for female offenders can receive points.

Program Fidelity

PFS 1A. When rating a program on its fidelity to the CDCR's risk assessment standards, the following scoring rules apply:

1. A program loses 4 points if there is a greater than 10% departure from risk assessment determinations in program assignment.
2. A program loses 3 points if it combines offender risk levels in the program groups.

12

2. Needs Assessment at Pre-Program Phase

"A program conducts or relies on a needs assessment instrument to determine services required by the offender. To receive points a program must conduct "meaningful" needs assessment. That is, needs identified by the assessment must align with a case plan. A program must target the criminogenic needs of the offender. A program must use an assessment tool that has been shown reliable and valid in previous research. A program that relies on needs assessment conducted elsewhere counts."

EIS 2A. "The needs principle," states: "Programs should address criminogenic needs." Criminogenic needs are the deficits each offender faces that have contributed to past offending behavior and are likely to lead to future offending unless they are addressed. These needs are dynamic risk factors, ones that it is possible to change through effective intervention.

In order to conform to the needs principle, three conditions must hold:
1. A needs assessment instrument must be used.
2. The needs assessment instrument must be valid.
3. The needs assessment must be meaningful.

A program must rely on an assessment of the needs of the offender, ideally one that presents an overall profile of all the offender's needs and prioritizes them according to the relative importance of addressing each one. It is important that the needs assessed and addressed be *criminogenic* needs, such as substance abuse, antisocial attitudes, low self-control, and dysfunctional family. Programs that address offender deficits that do not have a demonstrated relationship to criminal behavior, such as low self-esteem (for male offenders), may be valuable, but they are not change programs, and are a lower priority for the CDCR.

It is not necessary that the proposed program conduct its own needs assessment. It may rely on needs assessments done elsewhere.

Needs assessment instruments are valid if they have been demonstrated to effectively identify criminogenic needs, through a process of formal evaluation. Ideally, the needs assessment instrument will not only have been validated, but have been validated specifically for the CDCR offender population, or any subset of that population targeted by a program. Which needs are criminogenic, and the priority of addressing them, may vary across geographic populations, so the very best risk assessment instruments are those that have been tested on the specific program population, and failing that, one as like the program population as possible.

Conducting a needs assessment of a program population has little value if the results of that assessment are not applied in the operation of the program. A needs assessment is said to be meaningful if it has reliable consequences for program placement and case management. The program should take only participants who exhibit the criminogenic

need or needs that the program is designed to address, and the needs identified must match the program plan for that offender.

Key Questions
- What needs assessment instrument is being used on the program population? How will the program ensure that all participants have been assessed for criminogenic needs?
- Are the needs addressed by the program linked to criminal behavior?
- How does the program use the information provided by the needs assessment?

Scoring Rule

A program receives 4 points for utilizing a valid needs assessment and delivering program services accordingly. Points are awarded on an all or nothing basis. A needs instrument is considered valid if there has been any validation study of it within a five year period using California data. Tests of either predictive or construct validity are acceptable. A longer period can be accepted if the case can be made that the inmate or parole population has not fundamentally changed since the last validation. For programs serving female offenders, only use of needs assessment instruments that have been validated for female offenders can receive points.

Program Fidelity

PFS 2A. When rating a program on its fidelity to the CDCR's needs assessment standards, the following scoring rules apply:

- A program loses 4 points if there is a greater than 10% departure from needs assessment determinations in program assignment.

3. **Program Model**

Programs should have a clearly articulated rehabilitative model. A good program model has a number of attributes and the CPAP scoring rules are designed to capture and rate the ones with a demonstrated relationship to recidivism reduction. Scored elements are part of the program model if they are formally part of the program, contained in a written program manual, curriculum, or training materials. If elements are present as practices, but have not been formalized, points should not be awarded, because informal practices are much less likely to be consistently applied.

<u>Theoretical Model</u>

"The program must be based on a clearly articulated theoretical model that links the intervention content directly to an offender's criminogenic need"

EIS 3A. A program's theoretical model posits a cause and effect relationship between the program activities and a reduced likelihood or recidivism. The program model must identify a criminogenic need (for a partial list of recognized criminogenic needs, see the entry for "criminogenic needs" in the Glossary.), and it must link the program intervention to addressing that need. Note that female offenders have different criminogenic needs than male offenders.

<u>Scoring Rule</u>

A program receives 2 points for having a clearly articulated theoretical model linked to a criminogenic need. Programs do not receive points if they address only non-criminogenic needs (although they may address such needs alongside criminogenic needs), or if program content does not have a clear relationship to that need.

<u>Program Materials</u>

"Program manual or curriculum materials exist"

EIS 3B. Written program manuals and curricula are important means by which the program's theory and content is transmitted consistently to program staff. This is necessary for program rigor, consistency, accountability, and replication. If a program does not have these materials, it raises serious doubts as to the ability of the program to deliver on the promises of its theoretical model.

<u>Scoring Rule</u>

A program receives 2 points for having a program manual of curriculum materials. Points are awarded on an all or nothing basis.

<u>Cognitive Behavioral and Social Learning Methods</u>

"Uses cognitive behavior or social learning methods."

EIS 3C. Cognitive behavioral and social learning approaches have a track record of success in programs to reduce re-offending. Cognitive-behavioral theory posits that offending behavior is the result of patterns of thought that are conducive to criminal behavior, which are referred to by phrases such as "antisocial attitudes" and "criminal personality." Addressing these criminal thought patterns requires social learning techniques, in which the offenders are not only taught different ways of thinking, but model them in directed behavioral practice. Pro-social attitudes and behaviors are positively reinforced by program staff. Cognitive-behavioral and social learning methods stress the importance of structure, in which authoritative staff provide organized values, roles, rules and responsibilities; and of accountability, in which offender conformity to those values, rules and responsibilities are supported by appropriate positive and negative reinforcement.

Scoring Rule

A program receives 2 points for utilizing cognitive behavioral or social learning methods. Points are awarded on an all or nothing basis. Cognitive behavioral and social learning methods may not be effective for individuals with low mental functioning. This item may not be scored for programs that target such individuals.

Enhancing Intrinsic Motivation

"Program enhances intrinsic motivation."

EIS 3D. A degree of intrinsic motivation is necessary for an offender to realize lasting behavioral change. Offender motivation for change is likely to fluctuate over the course of program participation and the offender may experience substantial ambivalence about abandoning long-held patterns of thinking. Program staff can play a powerful supporting role in enhancing the motivation of inmates to change, using a technique called "motivational interviewing." Motivational interviewing is a directive, goal-oriented counseling style intended to elicit offender ambivalence about change in order to effectively resolve it.

Scoring Rule

A program receives 1 point for utilizing motivational interviewing techniques to enhance intrinsic offender motivation.

Continuities with Communities

"Program is structured to produce continuities between the program activities and communities, families, and other programs."

EIS 3E. Programs should produce continuities between program activities and pro-social offender support networks. Many successful program interventions recruit and use offender family

members, community programs and other sources of pro-social support to positively reinforce desirable behaviors. Engaging such support networks can extend the reach, and therefore the effectiveness, of programs in both time and space. It is important that the program differentiate between pro-social support networks and any community relationship network as family and peer relationships are a criminogenic risk for many offenders.

Scoring Rule

A program receives 1 point for producing continuities between program activities and communities and 1 point for continuities between program activities and families or other programs.

Intensity

> *"Program dosage varies with offender risk level (Higher risk offenders receive greater program dosage)"*

EIS 3F. Dosage refers to the total program exposure, generally measured in hours. The effectiveness of good programs can be diluted when the program is delivered at a low intensity. As a general principle, a higher dosage of programs should be delivered to higher risk offenders.

Intensity refers to how "compact" the program is, over how much time a program participant receives the program dosage. 100 hours of treatment over 3 months is a high dosage, high intensity program. 100 hours over the course of a year (or 2 hours a week) remains a high dosage program, but is fairly low intensity. A 50 hour program is not high dosage, but could be intensive over a short period, say 2 or 3 weeks. Thus, more program dosage can be delivered by either making the program more intense over the same period of time, or by increasing the duration of the program without diluting the intensity.

Intensity does not have to be the product of one program; multiple programs can combine to form an intensive intervention. The ideal for high-risk offenders is a high-intensity, high dosage program. Research has estimated the ideal dosage for community programs as 40-70% of a high-risk offender's free time over the first 3-9 months following release.

Scoring Rule

A program receives 1 point for delivering programs more intensively to higher-risk offenders. Programs that do not determine the risk level of participants cannot receive points for this item. For programs that only serve high-risk offenders, this item is not scored.

<u>Responsivity</u>

"Program design reflects the responsivity principle (i.e., it has procedures to determine the preparedness of the offender for the program and to match the delivery of the program to the learning style of the offender)"

EIS 3G. The "responsivity principle," states: "Programs should be responsive to the temperament, learning style, motivation and culture of offenders." These offender attributes can act in two ways important for program effectiveness. First, attributes such as offender motivation may determine whether an offender is "ready" for the program. An unready offender may be best excluded from a program, despite having the risk profile and criminogenic needs appropriate for participation. Second, once an offender is included in a program, the program will enhance its effectiveness by matching delivery to the different learning styles, temperaments and cultural backgrounds of the participants.

Taking all of these factors into account is a tall order for any program and a program that was able to account for all of them would be extremely impressive. Few programs are likely to do so. The CDCR does not expect change programs to be able to account for all of these factors, but does expect that programs will incorporate responsivity to relevant offender differences into their program model. "One size fits all" approaches do not offer the best chance of success.

<u>Scoring Rule</u>

A program receives 1 point if it is responsive to offender motivation, culture, temperament or learning style. No points are awarded if the program is delivered to all offenders in the same manner.

<u>Use of Positive Reinforcement</u>

"Program design identifies positive reinforcement strategies, not just sanctions"

EIS 3H. Programs should use positive incentives. The current research consensus is that positive reinforcement should be applied more heavily than negative reinforcement when trying to effect behavior change. A positive to negative reinforcement ratio of 4:1 is ideal. Programs should indicate their structure for employing positive and negative reinforcement, and better programs will incline towards the former. While sanctions must be reliably and consistently applied, research indicates that positive reinforcement does not need to be consistent in order to be effective.)

<u>Scoring Rule</u>

A program receives 1 point for formal use positive reinforcement.

Program Fidelity

PFS 2A. When rating a program on its fidelity to the CDCR's program model standards, the
following scoring rules apply:

- A program loses 6 points if implementation does not align with the program
  model.
- A program loses 2 points if continuities between in-prison, reentry and
  community supports are not well effectuated, defined as less than 50% of program
  participants linked across domains.
- A program loses 1 point if program intensity is not related to risk level in practice
- A program loses 1 point if responsivity assessment does not determine program
  delivery.
- A program loses 1 point if positive reinforcement is absent.

<div style="border:1px solid black; padding:10px">

Key Questions
- Is the theoretical underpinning of the program clear and persuasive?
- Does the program address a criminogenic need?
- Is the application of the model captured in written program material?
- Does program service delivery incorporate cognitive behavioral and social
  learning methods?
- Does the program enhance offender motivation to change?
- How does the program connect its activities with pro-social support networks
  in the offender's community?
- Is the intensity of program delivery higher for higher-risk offenders?
- How is the program responsive to differences in offender learning styles,
  motivation, culture, etc.?
- Does the program utilize positive reinforcement?

</div>

4. **Program Administration**

Program staff training and education should be appropriate and adequate for the purpose of the program. Even a well-designed program is unlikely to accomplish its goals if the staff of that program are not adequately qualified and trained.

Staff Education

> *"75% or more of service staff possess an undergraduate degree. Among those with degrees, 75% of staff has degrees in a helping profession"*

EIS 4A. CDCR prefers programs in which the staff directly engaged in the delivery of program services have undergraduate degrees and that the preponderance of those degrees be in a helping profession. Formal education is particularly important given that many of the elements of effective interventions that have proven effective (such as motivational interviewing and cognitive behavioral methods) have specialized technical content.

In particular, program staff members with degrees in helping professions, such as social work or psychiatry, are desirable. For programs with applicable content, medical or education degrees may be the most desirable.

Scoring rule

A program can receive 2 points for staff education. Programs receive 1 point if 75% of service staff has an undergraduate degree. The program receives another point if 75% of the staff with degrees has degrees in a helping profession.

Staff Experience Working with Offenders

> *"75% of staff has worked in offender treatment programs for at least two years"*

EIS 4B. Working with offenders to change their behavior presents unique challenges. Therefore, the CDCR prefers programs staffed primarily by people with substantial prior experience working with offenders.

Scoring Rule

A program receives 1 point if 75% of the staff has worked in offender treatment programs for at least two years.

Staff Recruitment and Retention

> *"Explicit strategy for recruitment and retention of staff"*

EIS 4C. Programs will be much more effective in recruiting a staff that meets CDCR's preferred standard if they have an explicit strategy for recruiting individuals with the desired

qualifications. A staff retention strategy to keep staff members in the program is also important. Heavy staff turnover interferes with the consistency of program delivery, and can cause deterioration in quality of even the best-designed programs.

Scoring Rule

A program receives 1 point if it has a strategy for recruitment and retention of staff. This may include provision of incentives such as opportunities for advancement and ongoing training.

Staff Training

"Initial training on program model, including written materials"

EIS 4D. Staff training is vital for the consistent delivery of program services in accordance with the program model. Written training materials best ensure the translation of the program model into program practice. Conversely, the absence of such material raises serious red flags regarding the quality of staff training.

Scoring Rule

A program receives 1 point if it provides training for staff that includes written training materials. Rating is based on the initial training provided to new staff members, not on in-service or refresher training.

Program Director Qualifications

"Program director was involved in the design of the program, has at least 3 years of experience with offenders, and has a degree in social work or related field."

EIS 4E. The qualifications and degree of involvement of a program director impacts the likelihood of program effectiveness significantly. The CDCR prefers program directors to have been involved in the development of the program, which provides them with greater knowledge of the program model, and for them to have experience working with offenders and a degree in social work or a related field. This familiarity with the program, experience in working with offenders, and professional education equip a program director to adjust and adapt the program once in the implementation phase.

Scoring Rule

A program can receive 3 points for program director qualifications. Programs receive 1 point if the program director was involved in the design of the program, 1 point if the director has 3 years of experience working with offenders, and 1 point if the director has a degree in social work or a related field.

Program Fidelity

PFS 4AB When rating a program on its fidelity to the CDCR's program administration standards, the following scoring rules apply:

- A program loses 2 points if the training or education requirements for staff are relaxed in practice.
- A program loses 3 points for high staff turnover, defined as more than 25% of staff in the course of a year.

Key Questions
- What percentage of the staff has at least a bachelor's degree? What percentage of those degrees are in a helping profession?
- Has the staff worked with offenders before? For how long?
- How does the program plan to recruit and retain staff?
- How does the program train its staff?
- What qualifications does the program director have? Was the director involved in the design of the program?

---

5. **Quality Assurance**

"Program collects data to monitor performance, includes individual level data on participation, identifies the eligible population, and data is forwarded and analyzed by a non-program entity"

Evidence-based practice requires not only evaluating evidence collected in other contexts when deciding what program approach to adopt, but collecting and using evidence once a program is in place. Programs should measure performance and use that information for continuous improvement. Effective measurement must be built into a program from the start in order to produce the most accurate and most useful data for program evaluation and improvement.

For this purpose, the CDCR looks for programs to do four things:
1. Collect data to monitor program performance.
2. Include individual level data on participation.
3. Identify the eligible population.
4. Forward data for analysis by a non-program entity.

Collect data to monitor program performance

EIS 5A. The program must have a mechanism to collect data on program performance. Collection of this data is a necessity in order to analyze program performance against the goals of the program, and apply that data to address any shortcomings. It should be clear who is collecting that data and how it will be aggregated.

Include individual level data on participation

EIS 5B. Collecting aggregate program data is insufficient for CDCR purposes. The CDCR needs individual level on program participation so that it can link program participation and outcomes to individual recidivism or parole success in the future, to better evaluate whether the programming available to offenders is realizing the CDCR mission of preventing further offending.

Identify the eligible population

EIS 5C. CDCR needs to know is how program participants do compared to similar offenders who do not participate in the program. This requires that the program clearly identify the eligible population for the program (offenders with a substance abuse problem, violent offenders, female parolees, etc.), so that a comparison of program participant outcomes to those of a group of non-participants within the eligible population is possible.

Forward data for analysis by a non-program entity

EIS 5D. Program data should be analyzed by a non-program entity, in order to ensure objectivity. A program can and should utilize its own data for monitoring participant progress and

making corrections and improvements in program delivery, but the analysis for the point of determining the program's effectiveness should be done by an outside entity, whether the CDCR Office of Research, academic, or other outside evaluation organization.

Scoring Rule

A program can receive 4 points for complying with CDCR's quality assurance standards. A program receives 1 point if it collects data to monitor performance, 1 point if it includes individual level data on participation, 1 point if it identifies the eligible population, and 1 point if it forwards data to a non-program entity for analysis.

Program Fidelity

PFS 5.   When rating a program on its fidelity to the CDCR's quality assurance standards, the following scoring rules apply:

- A program loses 2 points if it reports data inconsistently.
- A program loses 1 point for failure to adhere to eligibility requirements in program assignment.

> **Key Questions**
> - What data will the program collect? Who will collect it? Will that data allow an analysis to determine whether the program is realizing its goals?
> - Is data collected at the individual level?
> - What is the eligible population for the program?
> - What non-program entity will analyze the program data?

RESEARCH BASIS SCALE

Extent of Research Evidence

The *Research Basis* dimension of the CPAP is meant to appraise evidence around the effectiveness of a specific program model. This is done by examining program evaluations that demonstrate the promise to reduce recidivism among participants. This dimension is subdivided into two parts. The first part measures the <u>extent</u> to which the program model has been assessed or evaluated, and the second part measures the <u>quality</u> of existing evaluations.

<u>Expert Recommended</u>

"An expert committee, respected advisory group, or Best Practices panel recommends"
(1 point)

RBS 1. A program receives credit for this if it demonstrates that an outside committee, group or panel recommended it. The recommending entity should be comprised of experts in the relevant field (e.g., substance abuse treatment, mental health services) and should have based its recommendation on a systematic review of the program's implementation, function, and outcomes.

> EXAMPLE: Juvenile Justice Educational Enhancement Program (JJEEP) 2004 Annual Report. See, in particular, Chapter 10, "Case Studies of High Performing Programs."
>
> http://www.jjeep.org/annual2004/chapter10ar04.pdf (see p.157)

> EXAMPLE: "Youth Violence: A Report of the Surgeon General" (2001). See, in particular, Appendix 5-B, "Descriptions of Specific Programs that Meet Standards for Model and Promising Categories."
>
> http://www.surgeongeneral.gov/library/youthviolence/chapter5/appendix5b.html

> EXAMPLE: "In August 1998, Project PROVE earned a Best Practices for Education award from the American Correctional Association at the 128th Congress of Correction."
>
> http://www.people.vcu.edu/~vshivy/INTUIT/INTUIT%20summary.htm

"Multiple positive evaluations exist"
(2 points)

RBS 2. A program receives credit if it can identify more than one evaluation study that showed an association between program participation and recidivism reduction.

If multiple positive evaluations exist that are of Level 3 quality or higher, a program receives two points on this item. If multiple positive evaluations exist, but only one is of Level 3 quality, or none are, a program would only receive one point on this item.

"Published in peer reviewed outlet"
(2 points)

RBS 3. A program is credited if it can cite a *positive* evaluation in a peer reviewed publication. Such publications would typically be found in academic and practitioner journals. They would not include policy reports, self-published evaluations, or any other evaluation that was not screened by anonymous peer reviewers.

"Negative or no effect evaluations"
(-1 point)

RBS 4. If evaluations exist that show the program has no impact on recidivism, or increases recidivism, a point should be subtracted from the research basis score.

Quality of Evaluation Evidence

The research rigor component is based on the degree to which an evaluation has demonstrated that a program reduces the recidivism of its participants.[2] Specifically, this component assesses how well an evaluation answers the following question:

How much higher would program participant recidivism have been in the absence of the program?

The best evaluations adhere closely to a scientific evaluation model that measures the recidivism of program participants (the "treatment" group) against the recidivism of one or more *equivalent* groups of offenders ("comparison" or "control" groups). This is typically done through the identification of a sample of similar offenders who do not participate in the program, comparing the recidivism of program participants with the recidivism of the comparison group. The very best evaluation designs randomly assign offenders to the program and to a comparison group, which equalizes offender characteristics across groups and convinces readers that reductions in treatment group recidivism are due solely to program effects. Such evaluations meet the standards for Level 4.

While random assignment is the gold standard of program evaluation, many evaluations are not able to use this design because of ethical, legal and logistical barriers. Where random assignment is not possible, evaluations can use statistical methods to control for differences between groups. These evaluations fall under Levels 2 and 3, and are commonly called *quasi-experimental designs*. It is the degree to which evaluators demonstrate group comparability that determines the relative strength of these types of evaluation designs. When program and comparison groups are clearly comparable, readers may conclude with a high degree of confidence that reductions in recidivism are due to program effects. When groups do not seem comparable, reductions in recidivism may be due to program effects, or to other differences between the treatment and comparison groups, and it becomes difficult to discern whether program participation, or these other differences, are driving recidivism patterns.

The weakest program evaluations – Level 1 – lack any sort of similar comparison group. A comparison group may be present, but it would be so different from the program group that reductions in participant recidivism could be attributed to any number of factors, apart from the program. In other cases, there may be no comparison group at all and evaluators may base their conclusions on "before and after" effects, or simple statistical correlations between program participation and reduced offending. In each of these cases, the reader cannot be convinced that the program itself caused changes in offending behavior, as these changes may have resulted from a host of other factors.

---

[2] Studies done in other countries are also to be included in this process.

Each program's strongest evaluation study will be assessed using the research rigor scale. Points are awarded for positive evaluations that meet the standards for a particular level of rigor. A program, therefore, can receive anywhere from zero to ten points total in this part. This total is independent of the other four items on the research basis scale.

It is important to keep in mind that the distinctions between levels will not always be clear. A study may exhibit traits that situate it between levels and it will be up to raters to use informed judgment when awarding rigor points.

In the following sections, we provide more information about each level of research rigor, identify characteristics typical to studies in each level, and describe some sample evaluation studies that fall into each level.

Key Questions

When assessing the rigor of a particular study, keep the following questions in mind:

- How were the treatment and comparison groups assigned?
- How might the assignment process have biased group characteristics?
- What *observable* differences were there between treatment and comparison groups?
- What *unobservable* differences might there have been between treatment and comparison groups?
- What steps did evaluators take to control for observable and unobservable group differences?
- Did evaluators convincingly demonstrate that treatment group subjects would have re-offended more in the absence of the program?

Level 1: Correlation between program participation and recidivism reduction , temporal sequence between program participation and recidivism reduction clearly observed, or a comparison group present without demonstrated comparability to the treatment group (& no controls) (1 point)

RBS 5A. Level 1 evaluations are relatively weak, though some evidence of program effectiveness will be presented. Level 1 studies are seldom found in reputable academic or practitioner journals, as they will not exhibit the social scientific rigor that such publications demand. More often, these studies will be found in the form of reports, including self-evaluations. They may also be found in journalistic outlets like newspapers and magazines. Results may be presented anecdotally.

These studies may only demonstrate an association between program participation and reduced offending, either through statistical correlations between participation and reduced recidivism, or simply the presentation of a low post-program recidivism rate. Multivariate statistical approaches may not be used at all, and if they are, statistical controls do not capture many of the critical differences between groups. As a result, reductions in the recidivism of program participants could be attributed to the program itself, or other cross-group differences, and it will be impossible to disentangle these effects.

Other studies in this group may compare program participants' recidivism against the recidivism of comparison groups that are not similar to the program group. No attempt is made to match the program group to comparable subjects in other groups. This makes it impossible to tell whether recidivism reduction was due to the program, or factors outside of the program.

Some evaluations in this category may use a *one-group pretest-posttest* design (or *temporal sequence design*), which compares treatment subjects' post-program recidivism to pre-program offending patterns. By demonstrating reductions in offending after treatment, evaluators hope to suggest a program effect. This approach is flawed, however, in that reductions in offending may be due to treatment effects, or something else. Without simultaneous examination of a comparison sample under a different condition, it is very difficult to separate the effect of a program from the effects of court involvement, subject characteristics, and other relevant factors. Remember that our critical question of interest is: *How much higher would program participant recidivism have been in the absence of the program?* Studies using temporal sequence designs have a hard time answering this question.

LEVEL 1 EXAMPLE: Dominguez (2005). 2nd Chances: A Novel Prison Program in New York City uses Nature to Teach Inmates about Life's Larger Lessons. Audubon Magazine.

This is an Audubon Magazine article about Project GreenHouse, a gardening program for inmates on Riker's Island in New York City.

As this is a news piece, there is no social scientific approach to evaluation. The main indication of outcome is the following passage: "For Riker's inmates, the recidivism rate—the percentage of them who end up back in prison—is thought to be about 65 percent. For those in the GreenHouse Program, the rate is 10 percent."

This type of evidence of program effectiveness meets only the barest minimum requirement for research rigor. Simply stating the recidivism rate of participants and comparing it to a general figure provides no indication of whether these offenders would have recidivated in the absence of the program. Since there is no source associated with the 65% figure, the comparison being made in this article cannot be validated. How would GreenHouse participants have fared in the absence of the program? Offenders in the GreenHouse program may have been less likely to reoffend than those in the larger Riker's population anyway, and this could have explained the low recidivism rate. The question of program effectiveness therefore remains open.

http://magazine.audubon.org/features0505/rehabilitation.html

LEVEL 1 EXAMPLE: Carney and Buttell (2004). A Multidimensional Evaluation of a Treatment Program for Female Batterers: A Pilot Study. Research on Social Work Practice 14(4): 249-258.

The authors conducted a preliminary evaluation of a batterer intervention program for women by investigating changes in psychological variables related to abuse (i.e., truthfulness, violence, lethality, control, and stress coping abilities) between pretreatment and posttreatment assessments in a sample of women involuntarily placed in treatment. This study evaluated arrest records for a period of 12 months following treatment completion to determine the association between changes on these psychological variables and recidivism.

Multiple problems plague the study design. First, the evaluators only examine the recidivism outcomes of program completers (26 women out of 139), resulting in a very small study sample. No other comparison group is present. The omission of program non-completers from the analysis obscures the interpretation of results, as completers may possess distinct characteristics associated with both program completion and a low risk of recidivism. This design cannot identify such characteristics, nor can it separate the effects of those characteristics on recidivism from the effects of the program.

Only one of these twenty-six women was rearrested within 12 months, so no statistical analyses were possible. How many would have been rearrested in the absence of the program? And did the program have any benefit for those women that did not complete it? In the end, the recidivism of treated women was low (1/26), but it was impossible to determine whether the program itself was responsible for this outcome.

LEVEL 1 EXAMPLE (program completion correlated with lower recidivism): Murphy, Musser, and Maton (1998). Coordinated Community Intervention for Domestic Abusers: Intervention System Involvement and Criminal Recidivism. Journal of Family Violence 13(3): 263-284.

This is an example of demonstrating program effectiveness by correlating involvement with reduced recidivism. The program was a "coordinated community intervention approach" involving court actors and community-based service providers in Baltimore. The evaluators studied male perpetrators of domestic violence (n=235), who were referred to the Maryland state attorney's Domestic Violence Unit. Different things happened to these subjects. Some were found guilty. Others received deferred judgment. Other cases were dropped. Some received probation while others got suspended sentences. Some were ordered to domestic violence counseling.

No comparison group was present. The claim to program effectiveness came from the finding that deeper program involvement was associated with decreased recidivism. Specifically, subjects who were ordered to domestic violence counseling had lower recidivism rates than those who were not. Among subjects ordered to counseling, those who completed the domestic violence counseling intake process had lower recidivism than those who had not. Finally, those completing the counseling program exhibited the lowest recidivism of all.

The study's findings regarding the promise of this approach in reducing recidivism are suspect. Most results are presented as descriptive statistics that simply report the recidivism rates of subjects in various system statuses. There is a multivariate model predicting recidivism, but it only controls for age, legal representation, history of violence, and the severity of injury related to the original offense. While this model does produce a significant coefficient for "system involvement," it is difficult to draw meaningful conclusions because so many potential explanatory factors are excluded.

The main problem with the findings has to do with selection bias. Subjects who demonstrated the deepest program involvement may be those who were most highly motivated for treatment. That is, those who completed the counseling intake were likely to be more amenable to counseling than those who did not complete intake. Similarly, subjects who completed the counseling program may have been more "treatable" than those who failed to complete the program. Therefore, it is no surprise that deeper involvement was found to be associated with lower reoffending rates. Without any measures to control for the range of factors that are likely to be correlated with both program involvement and the risk of recidivism, the dynamics of treatment, motivation and recidivism are obscured.

LEVEL 1 EXAMPLE (non-comparable controls): Fielding, Tye, Ogawa, Imam and Long (2002). Los Angeles County Drug Court Programs: Initial Results. Journal of Substance Abuse Treatment 23(3): 217-224.

This study attempted to evaluate the effectiveness of drug courts in Los Angeles, comparing the recidivism of drug court participants to the recidivism of two comparison samples:

> 1. Participants in a drug diversion education program.
> 2. Felony drug offenders in neither program.

Descriptive statistics suggested serious differences between groups on demographic, social and legal measures. There were even differences in the distribution of assessed risk scores across the groups. Selection into treatment and control groups was not random, and there is reason to believe that comparison subjects, particularly those in neither program (comparison sample 2), were less amenable to treatment and more prone to re-offending. In fact, the riskiest offenders were not even permitted to participate in the drug court. Therefore, all of these risky subjects fell into the two comparison samples.

No multivariate analysis was attempted. The evaluators simply compared re-arrest rates and average times to re-arrest across groups. While they did find drug court participants to have lower levels of recidivism than the comparison samples, this finding may very well have been the result of preexisting group differences. There was no way of separating program effects from the effects of other offender characteristics.

LEVEL 1 EXAMPLE (temporal sequence design): Lee, Uken and Sebold (2004). Accountability for Change: Solution-Focused Treatment with Domestic Violence Offenders. Families in Society 85(4): 463-476.

This study evaluates a "solution-focused treatment program for domestic violence offenders." According to the article, this approach holds a person accountable for solutions instead of focusing on problems. Study subjects were 90 male and female domestic violence offenders ordered to the program in lieu of prosecution.

Evaluators used a one-group pretest-posttest design to gauge changes that may have resulted from the program. The principal focus of the analysis was monitoring changes in relational dynamics between offenders and their partners, and researchers applied a series of clinical instruments to accomplish this task. Recidivism was a secondary focus, captured both by official sources (post-treatment only) and self-reports. For self-reports, offenders and their partners were interviewed six months after group participation. They were asked about the extent of domestic violence before the program, and after it.

Official records indicated an overall recidivism rate of 16.7%, which according to the authors, "was considerably lower than that for most other conventional treatment programs." This finding is weak, as there is no clear comparison being made. Offenders in other treatment programs may have been very different from those in the current study. Further, it is impossible to assess how well these study subjects would have done in other settings, or how offenders in other settings would have done in this program.

Self-reported recidivism measures allowed some measurement of a program effect, although this approach, too, was severely flawed. Study subjects and their partners reported lower post-program rates of verbal and physical domestic violence. However, with no comparison sample present, it was not possible to say whether this decrease in violence was due to treatment, or other factors. For example, contact with the justice system alone may have accounted for this effect. Subjects may have been deterred by the court experience from committing further violent acts, and the treatment program itself may have had no benefit whatsoever. Whatever the truth, this study design could not fully measure this program's effect. At best, it served as a suggestion of effectiveness.

Level 2: A comparison between two or more units of analysis, one with and one without the program (with partial controls) (4 points)

RBS 5B. Level 2 evaluations compare the recidivism outcomes of treatment and control samples, and like Level 1 studies, they are limited by cross-group differences. However, Level 3 evaluations are superior to Level 1 evaluations in that they make some attempt to control for these differences. This is typically done through the inclusion of important observed independent controls (e.g., gender, age, offense type, criminal history) in multivariate statistical models. Independent variables effectively control out the impact of these factors on recidivism outcomes, thereby better isolating the unique effect of the program itself.

The inclusion of independent controls basically renders study groups equal on these *observed* measures. However, what these types of designs cannot control for are *unobserved* differences between groups. That is, treatment and control groups may also differ on "invisible" factors such as inherent criminality, motivation for treatment, emotional stability, and anything else that independent measures do not, or cannot, capture. The danger is that these unobserved group differences may also be affecting recidivism, and obscuring the measured effects of treatment.

The selection process into treatment and control groups is often the main culprit behind group differences. When random assignment is not possible, there must be some mechanism for assignment to, and exclusion from, the program being studied. The process of sorting program participants from non-participants can produce a treatment sample that is fundamentally different from non-treatment samples. Say, for example, that participation in a residential drug treatment program is voluntary. Those who sign up for the program are ostensibly more motivated for treatment than those who do not choose to enroll. If program participants exhibit better recidivism outcomes than comparison subjects, this underlying difference in motivation may be the critical explanatory factor. The program itself may only be exerting a minimal effect on recidivism, but will appear to be quite effective. If evaluators cannot control for everything that sorts offenders into and out of the treatment group, they cannot say for certain that the program is behind changes in offending patterns. *When assessing the strength of an evaluation study, pay very close attention to the means by which subjects were selected for treatment.*

The point is that in Level 2 studies, unobserved differences may be behind differential group outcomes, but the evaluator cannot know this. This may result in false conclusions of program effectiveness (or ineffectiveness). Having controlled for observed group differences, evaluators may attribute recidivism reductions to program participation, but in truth, unobserved group differences will be the true cause of differential recidivism outcomes. Or, the program may have some benefit, but this benefit will be overstated, as unobserved group differences will partially account for this effect.

LEVEL 2 EXAMPLE: Rodriguez and Webb (2004). Multiple measures of juvenile drug court effectiveness: Results of a quasi-experimental design. Crime and Delinquency 50(2): 292-314.

The authors reported on a 3-year evaluation that examines the impact of juvenile drug court participation on recidivism and drug use. A quasi-experimental design was used to compare juveniles assigned to drug court with those assigned to standard probation in Maricopa County, Arizona. Subjects were initially randomly assigned to drug court and to standard probation, but judges abandoned this plan a few months into the study because they felt they were denying treatment to needy juveniles. As a result, selection into treatment and control groups was vulnerable to bias due to judicial discretion.

Analyses of drug use showed no significant pretreatment difference between groups in marijuana use but revealed that drug court participants were more likely than juveniles in the comparison group to test positive for cocaine. To suppress the impact of cross-group differences, the evaluators statistically predicted group assignment and included significant variables in subsequent multivariate recidivism models. This was a step in the right direction, but it could not capture all differences between groups – particularly unobserved differences.

Findings indicated that drug court participants were less likely to recidivate than youths in the comparison group. These findings were convincing, as the drug court sample seemed "harder" (more needy) than the comparison group at baseline. Therefore, in the absence of treatment, drug court subjects would be expected to exhibit *higher* recidivism than controls. However, without a truly equivalent comparison sample, the real relationships between needs, treatment and outcomes could not be discerned with 100% accuracy.

LEVEL 2 EXAMPLE: Hall, Prendergast, Wellisch, Patten and Cao (2004). Treating Drug-Abusing Women Prisoners: An Outcomes Evaluation of the Forever Free Program. The Prison Journal 84(1): 81.

This article was an evaluation of the Forever Free program, which consisted of an intensive 6-month drug treatment program for female inmates near the end of their prison sentence, followed by community-based residential treatment for those women who volunteered for continued services.

The treatment sample consisted of 119 women who participated in the prison-based treatment component. Four women did not graduate, and only 47 chose to take part in the post-release treatment component. All 119 were included in the treatment sample, however. The comparison sample consisted of 96 women who voluntarily participated in a low-intensity substance abuse education program in prison. These women were selected as comparison subjects because of their apparent similarity to Forever Free clients.

Bivariate descriptive statistics showed that the study samples were indeed quite similar, though comparison group subjects were more likely to have received drug treatment during a prior incarceration, and less likely to have ever injected drugs. Bivariate recidivism analyses showed Forever Free clients had fewer rearrests and reconvictions than comparison subjects during the first postrelease year. In multivariate survival analysis models, recidivism was operationalized as reincarceration. Survival models suggested a positive treatment effect for Forever Free program participants, though this coefficient did not reach statistical significance.

This study is a good example of a quasi-experiment with potential selection bias. While bivariate and multivariate analyses showed evidence of a treatment effect, it is impossible to separate this effect from the potential effects of cross-group differences. Specifically, those women who volunteered for the Forever Free program may have been more motivated for treatment than women in the comparison group. This possible difference in motivation, along with other unmeasured differences, potentially explains some disparity in recidivism outcomes.

http://tpj.sagepub.com/cgi/reprint/84/1/81

LEVEL 2 EXAMPLE: Evaluation of the Indianapolis Violence Reduction Partnership (IVRP), which uses a problem-solving approach to reduce recidivism among former inmates (parolees and probationers).

This program for parolees and probationers in Indianapolis used a deterrence model that consisted of meetings with law enforcement officials, service providers, and community members who detailed the consequences of violent actions while under community correctional supervision.

For this study, the IVRP treatment group was mandated to attend by their probation and parole officers. The treatment sample was recently released from prison, and lived in three geographical areas that were of concern to the police. The comparison sample was comprised of inmates released to those same neighborhoods during the same time period. The process of selection into treatment and control samples is not well-documented, and there is a serious concern about differences between groups. Initial descriptive statistics showed observed differences on offense histories and demographics, though these were controlled for in multivariate models. The real concern here is around unobserved differences, as the selection process appeared far from random. Even with these concerns, the evaluation showed very little benefit from the program.

http://www.ncjrs.gov/pdffiles1/nij/grants/203923.pdf

Level 3: A comparison between multiple units with and without the program, controlling for other factors, or a nonequivalent comparison group has only minor differences evident (6 points)

RBS 5C. Level 3 evaluations, like those in Level 2, attempt to equalize treatment and control groups in their study designs. However, Level 3 studies take extra measures to achieve this equalization. In particular, these better evaluations make some attempt to control out unobserved differences, as well as those that are observable. Better Level 3 evaluations will approach random assignment by demonstrating convincing comparability between study groups.

Here are a few approaches that Level 3 evaluations commonly use:

1. **Matched samples**: Evaluators may match the treatment sample to a control sample. That is, they use a technique to find cases in the control population that are most similar to those in the treatment group based on important criterion variables. This method does a good job of equating study groups on observed variables, but still suffers from the possibility that groups are different on unobserved measures, especially if treatment group subjects volunteered for treatment. (They may have been more motivated.) However, by matching on observed factors, evaluators hope that observed cross-group similarities "soak up" unobserved differences as well.

2. **Propensity score matching**: This is similar to the matched samples approach, but instead of being matched on multiple criterion variables, subjects are matched on their predicted likelihood of inclusion in the treatment sample. A regression is run to identify independent predictors of inclusion in the treatment sample, and each subject is then assigned a propensity score for treatment group inclusion, based on this set of predictors. Treatment subjects are matched to control subjects on propensity score. Treated subjects and their matched counterparts are then compared on the outcome measure (recidivism) in order to determine the impact of treatment.

3. **Heckman two-step process**: Evaluators may try to control for group differences using the Heckman approach, which first models the likelihood of inclusion in the treatment group based on observed measures (often using a probit analysis), uses residuals from the selection equation to create a correction term for unobserved variation (or "selection bias control factor"),[3] then introduces this term as an independent variable in substantive multivariate models predicting recidivism. The correction term is meant to soak up unobserved cross-group variation, thereby isolating a true treatment effect.

4. **Regression-discontinuity design**: The best Level 3 studies use regression-discontinuity. This is a very strong method that uses no real comparison group at

---

[3] Specifically, this is the inverse Mill's ratio.

all. Basically, this design imposes a cutoff point for group membership – typically on a risk score. That is, offenders are scored using a standardized risk instrument, and only those above a certain score will participate in the program. Those who score below the cutoff are assigned to the "control" group – even though they are less risky than the treatment group. Then, for all subjects, a graph is created that maps pre-program risk against post-program recidivism. A regression line is fit to this "map." If the program has an effect, a jump will be observed in the regression line at the cutoff point. This jump will represent the magnitude of the effect of the program. Since these designs rely on clear separation of treatment and non-treatment samples at the cutoff point, any "leakage" across the cutoff will weaken results.

LEVEL 3 EXAMPLE (matched samples): Welsh (2002). Evaluation of Prison Based Drug Treatment in Pennsylvania: A Research Collaboration Between the Pennsylvania Department of Corrections and the Center for Public Policy at Temple University, Final Report. Washington, DC: National Institute of Justice.

This was an evaluation of a therapeutic community for drug users in Pennsylvania prisons. The treatment group went to the therapeutic community; comparison group members were attending less intensive types of drug treatment in the same facilities. The evaluator used a matched-sample design to control out observable differences between groups. Treatment and comparison samples were matched on demographics, drug dependency, need for treatment, and criminal history, so that distributions of these criterion measures were equivalent within each group. Formal assessment instruments were used, when appropriate, to score subjects for matching.

While matching seemed effective at equalizing study groups on observed measures, unobserved cross-group variation may still have been a problem. The evaluator does not talk about this much, but it appeared that participation in therapeutic communities was voluntary. This self-selection process may have hidden some unobserved differences between treatment and control groups. Matching may have erased some of these differences, but perhaps not all of them.

Results indicated reduced recidivism for treatment group.

http://www.ncjrs.gov/pdffiles1/nij/grants/197058.pdf (full report)
http://www.ncjrs.gov/App/Publications/abstract.aspx?ID=197058 (abstract)

LEVEL 3 EXAMPLE (propensity score matching): Rempel, Fox-Kralstein, Cissner, Cohen, Labriola, Farole, Bader, and Magnani (2003). The New York State Adult Drug Court Evaluation: Policies, Participants and Impacts. New York: Center for Court Innovation.

The evaluators assessed the effectiveness of six adult drug courts in New York State. Treatment subjects who were adjudicated in drug court were compared to control subjects who would have been eligible for drug court, were convicted, but for administrative or logistical reasons, did not go to drug court. Because differences between samples may also have been based on unobserved factors, the comparison sample was then further refined using propensity score matching. A logistic regression was run to generate propensity scores (predicting drug court sample inclusion) for all potential study subjects. Drug court subjects were then matched to comparison subjects with the nearest possible propensity score.

Bivariate comparisons of recidivism and multivariate recidivism models indicated reductions in recidivism for drug court subjects compared to subjects who had received conventional case processing. While these findings are persuasive, the possibility exists that unobserved differences between study samples explain some part of the treatment effect. As the authors themselves note, comparison group subjects may have been ineligible for drug court because of unrecorded charge characteristics (e.g., they were involved in heavy trafficking), because they were not addicted to drugs, or they may have simply refused to participate. These differences, which would clearly affect possible group membership, might also have some effect on the likelihood of recidivism.

Overall, this is a strong study, but it is not completely immune to risks associated with quasi-experimental designs.

http://nycourts.gov/whatsnew/pdf/NYSAdultDrugCourtEvaluation.pdf

LEVEL 3 EXAMPLE (Heckman two-step process): Harrell, Mitchell, Merrill, and Marlowe (2003). Evaluation of the "Breaking the Cycle." Washington, DC: Urban Institute.

Breaking the Cycle (BTC) is a "systemwide" intervention program that aims to reduce drug use and criminality among drug-involved felony defendants. Eligible defendants were ordered to report to BTC for drug screening as a condition of pre-trial release. Those who reported drug use, tested positive for drugs, or were arrested on drug felony charges were placed in drug testing and, when appropriate, referred to drug treatment or drug education classes. The key system reforms instituted under the BTC model were early intervention, judicial oversight, graduated sanctions and incentives, and collaboration among justice and treatment agencies.

Sites selected for evaluation were Birmingham, Jacksonville and Tacoma. The evaluation was based on a quasi-experimental comparison of defendants in BTC with samples of similar defendants arrested in the year before BTC implementation.

Treatment group members were not mandated to participate in the study. They were asked to participate, and some potential subjects chose not to. This self-selection process may have resulted in selection bias on unobserved characteristics. That is, those that chose to participate may have been different from those that chose not to participate.

Bivariate descriptive statistics showed inter-group differences across a variety of observed measures. These were controlled for in multivariate models, but unobserved group differences may still have affected the analysis. In multivariate models, authors tried to account for observed and unobserved differences by introducing a *Heckman correction term* (inverse Mill's ratio). Evaluators hoped that this adjustment would control for unobserved cross-group differences. While not a perfect solution, it was substantially better than simply including independent variables in regressions. Evaluators found that the program did serve to reduce recidivism.

Appendix A details sample selection and statistical approach.

http://www.ncjrs.gov/pdffiles1/nij/grants/205002.pdf
http://www.ncjrs.gov/App/Publications/abstract.aspx?ID=205002 (abstract)

LEVEL 3 EXAMPLE (regression-discontinuity): Berk and Rauma (1983). Capitalizing on Nonrandom Assignment to Treatments: A Regression-Discontinuity Evaluation of a Crime-Control Program. Journal of the American Statistical Association 78(381): 21-27.

These authors evaluated a California program that provided unemployment benefits to prisoners after release. The execution of the program provided a good opportunity to utilize a regression-discontinuity evaluation design, as only ex-inmates who had worked a certain number of hours in prison were eligible for the benefits. The evaluators compared eligible ex-inmates with ex-inmates who applied for the program but were denied because they had not worked enough hours. Desire to participate in the program was therefore not a source of selection bias, and "hours worked" served as a clean cutoff criterion for the regression-discontinuity design.

Descriptive statistics indicated that study subjects were very similar to the larger parole population. Results from multivariate regression-discontinuity modeling indicated statistically significant associations between program participation and reduced recidivism. Tests for model misspecification suggested a strong and valid analysis.

Level 4: Random assignment and analysis of comparable units to program and comparison groups (10 points)

RBS 5D. As stated above, the very best evaluations use random assignment to treatment and control groups. This simple method is nearly foolproof in that treatment and control samples should be equal on observed and unobserved characteristics, as there is no bias in the selection process. Any particular subject has just as much likelihood of receiving treatment as any other subject, regardless of their characteristics. Observed differences between groups (these should be the product of chance) can still be documented and controlled for in multivariate statistical models, but unobserved differences should be suppressed and findings regarding program impact should therefore be "purified."

While random assignment is ideal for program evaluation, such designs are relatively rare because of ethical, legal and logistical concerns. However, when such a design is used, findings should be taken very seriously.

When assessing a study that uses random assignment, it is important to confirm that selection into treatment and control groups was indeed random, and not corrupted in any way. If there is some discrepancy in the assignment process, examine it carefully, and determine whether this discrepancy might have affected recidivism outcomes. Random assignment studies that are not actually random may fall into Level 3, or even Level 2.

LEVEL 4 EXAMPLE (random assignment): Taylor, Davis and Maxwell (2001). The Effects of a Group Batterer Treatment Program: A Randomized Experiment in Brooklyn. Justice Quarterly 18(1): 171-201.

This is an evaluation of a domestic violence program in Brooklyn for male batterers. The authors conducted an experimental evaluation in which 376 adult males convicted of domestic violence were randomly assigned to either a 40-hour batterer treatment program or 40 hours of community service that did not include any therapeutic treatment. Random assignment was slightly compromised by cases in which judges overrode random assignment and sent the subject to the program anyway. However, the authors performed various statistical analyses (e.g., conducting a regression to predict overrides) to correct for this problem, and claim that their results remain unbiased.

Those assigned to the treatment program showed significantly lower recidivism on the basis of all outcome measures from official records. Although victims' reports also recorded fewer failures among the batterers assigned to the treatment group, the differences in failure rates were not large enough to be statistically significant. Overall results suggest that therapeutic treatment for batterers may reduce domestic violence among convicted batterers who agree to this sentence.

LEVEL 4 EXAMPLE (random assignment): Van Voorhis, Spruance, Ritchey, Listwan, and Seabrook (2004). The Georgia Cognitive Skills Experiment: A Replication of Reasoning and Rehabilitation. Criminal Justice and Behavior 31(3): 282-305,

This evaluation measured the effectiveness of the Georgia Cognitive Skills Program, a mandated cognitive therapy program for parolees. Parolees (n=468) were randomly assigned to treatment and comparison groups between May 1997 and July 1998. Descriptive statistics confirmed that the study samples were similar on observed measures.

Using survival analysis, the evaluation tested treatment effects on arrests/revocations, technical violations, and employment at 9 months, and returns to prison at 18 to 30 months. Evaluators found slightly lower (statistically insignificant) recidivism rates for experimental participants than comparisons. No significant differences were found between experimental and comparison participants on technical violations and employment. Statistically controlling for offender risk factors, program completers had significantly fewer rearrests/revocations and returns to prison and more favorable employment outcomes than comparisons and dropouts.

Because of the experimental study design, these findings are highly reliable. Observed cross-group differences were clearly a non-factor in skewing outcomes, and unobserved differences were not likely to have corrupted results either.

Tab 11: Glossary

Activity programs: Programs that occupy the time of the ward/offender, stimulate them, or engage them in pro-social activity to ease their adjustment to custody or to the outside.

Change programs: Programs intended to change the antisocial behavior of offenders/wards through rehabilitation, education, and employment programs.

Cognitive behavioral therapy: A highly structured psychotherapeutic method used to alter distorted attitudes and problem behavior by identifying and replacing negative inaccurate thoughts and changing the rewards for behaviors

*Comparison group***:** Those who do not receive the "treatment" (i.e., program) under consideration. Also, *control group*, *non-treatment group*.

*Comparison sample***:** Those in the comparison group who are included in the study. Also, *control sample, non-treatment sample*.

Construct validity: Evidence of validity gained by showing the relationship(s) between a theoretical construct and tests that propose to measure the construct.

*Control group***:** See *comparison group.*

Control programs: Programs designed to increase offender/ward compliance and enhance the supervision capabilities of correctional officials

*Control sample***:** See *comparison sample*.

*Correlation***:** A statistical association between variables that does not control for other relevant factors. Correlations do not imply a causal sequence.

Criminogenic need: Attributes of offenders that are directly linked to criminal behavior. Demonstrated criminogenic needs for male offenders include, but are not limited to:
- Anti-social attitudes, beliefs and values
- Criminal thinking
- Poor decision-making/problem-solving skills
- Low levels of educational/vocational achievement
- Poor self-control/self-regulation
- Substance abuse

According to research supported by NIC, criminogenic needs of female offenders may include the following:
- Housing safety

- Mental health (anxiety and psychosis)
- Childhood and adult abuse (physical and sexual)
- Substance abuse
- Parenting (parental stress)
- Self-efficacy
- Self-esteem
- Relationships (dysfunction)
- Family conflict

Dosage: The total amount of program exposure for an individual participant, generally measured in hours.

Dynamic risk factors: Offender attributes that are linked to criminal behavior and that can change or be changed.

Evidence-based practice: The conscientious, explicit, and judicious use by correctional administrators of current best research evidence in selecting programs designed to manage offenders, reduce recidivism, and increase public safety. Research evidence of program effectiveness must adhere to accepted methodological standards. A program must also use empirical assessment tools to target the program to the individual offender and must objectively measure program implementation and outcomes. Evidence-based programs also includes programs that adhere to 'principles of effective intervention' established by prior research.

*Experimental design***:** See *random assignment*.

Halo effects: The effects that one characteristic of an individual has on the rating of that individual in other respects. For example, if an individual is tall, the rater may perceive that individual to be more intelligent than he actually is.

*Heckman 2-step process***:** A quasi-experimental approach that first models the likelihood of inclusion in the treatment group, then constructs a correction term for unobserved cross-group variation using residuals from this first model. The correction term is then included in substantive models predicting recidivism to control for unobserved differences between groups.

Intensity: The "compactness" of a program, or the period of time over which program participants receive the program dosage, generally measured in months.

*Matched samples***:** A quasi-experimental approach that matches treatment and comparison groups on observable measures.

Motivational interviewing: A directive, client-centered counseling style for eliciting behavior change by helping clients to explore and resolve ambivalence about change.

Compared with nondirective counseling, it is more focusing and goal-directed. The examination and resolution of ambivalence is its central purpose, and the counselor is intentionally directive in pursuing this goal.

Multivariate modeling: A statistical approach that predicts program outcomes (recidivism), controlling for other important factors. Regression and survival analysis are common forms. When done well, multivariate models can claim a causal association.

Needs principle: Target correctional program interventions to criminogenic needs.

Negative evaluation: A program evaluation that demonstrates an association between treatment and increased recidivism.

No effect evaluation: A program evaluation that shows treatment to have no impact on recidivism.

Non-treatment group: See *comparison group*.

Non-treatment sample: See *comparison sample*.

Observed variation (or differences): Differences in group characteristics that are captured by observable indices. These differences can usually be controlled for in multivariate modeling.

One-group pretest-posttest design: A study design that lacks a comparison group. Treatment group offending is measured before and after program participation to estimate program effects. This is a relatively weak approach. Also, *temporal sequence design*.

Peer-reviewed outlet: A journal or other publication which only publishes articles that have been reviewed and approved by professionals in the relevant field. Reviews are typically done anonymously.

Positive evaluation: A program evaluation that demonstrates an association between treatment and decreased recidivism.

Predictive validity: The extent to which the predictive instrument is successful in anticipating the relevant outcome.

Propensity score matching: A quasi-experimental approach that first predicts the likelihood of membership in the treatment group, then matches treatment and comparison subjects on this likelihood.

***Quasi-experimental design*:** A study design in which treatment and comparison groups are not equal on observed and/or unobserved measures. Such designs may be limited by selection bias.

***Random assignment*:** The process of equalizing treatment and comparison samples by arbitrarily assigning subjects to each condition. This is the optimal design for program evaluation studies. Also, *experimental design*.

***Recidivism*:** Subsequent criminal activity. Re-offending.

***Regression-discontinuity design*:** A strong approach that uses no comparison group at all. This approach can only be used if treatment assignment is based on a clear scored criterion with a cutoff point for treatment group membership. Those on one side of the cutoff receive treatment and those on the other side do not. For all subjects, a graph is created that maps pre-program risk against post-program recidivism. A regression line is fit to this "map." If the program has an effect, a jump will be observed in the regression line at the cutoff point. This jump will represent the magnitude of the effect of the program.

Responsivity principle: Be responsive to the temperament, learning style, motivation, gender, and culture when assigning to correctional programs.

Risk principle: Prioritize correctional supervision and treatment resources for higher risk offenders.

***Selection bias*:** A problem common to quasi-experimental designs. When assignment to treatment and comparison groups is not random, observed and unobserved differences between groups can obscure the real effect of treatment.

Social learning theory: Theory positing that people can learn new behaviors, attitudes and feelings by observing other people and events, followed by individual practice of appropriate thoughts and behaviors.

Static risk factors: Offender attributes that are linked to criminal behavior and are inherent and cannot be changed.

***Subject*:** A person who is included in the study.

***Temporal sequence design*:** See *one-group pretest-posttest design*.

***Treatment effect*:** The measured impact of a program or service (in our case, on recidivism).

***Treatment group*:** Those who are to receive the "treatment" (i.e., program) under consideration.

*Treatment sample***:** Those in the treatment group who are included in the study.

*Unobserved variation (or differences)***:** Differences in group characteristics that are not captured by observable indices. These differences are the biggest problem in non-experimental evaluation research, as it is very difficult to control for them in statistical models.

Validated instrument: An assessment instrument that has been tested on the population of interest, and found to be predictive of an outcome of interest, such as likelihood of recidivism.

Tab 12: Selected Reference Sources

<u>Principles of Effective Change Programs</u>

Bonta (2000). "Offender Rehabilitation: From Research to Practice." Paper presented at the 2000 APAI Conference, Ottawa. Available at http://www.apaintl.org/Pub-Conf2000-PlenaryAndrews-En.html.

Gornik (2001). "Moving from Correctional Program to Correctional Strategy: Using Proven Practices to Change Criminal Behavior." National Institute of Corrections. Available at http://nicic.org/Library/017624.

Joplin, Bogues, et al (2004). "Using an Integrated Model to Implement Evidence-Based Practices in Corrections. Crime and Justice Institute/National Institute of Corrections. Available at http://crjustice.org/cji/NICCJI_Project_ICCA.pdf.

Lowenkamp and Latessa (2006). "The Risk Principle in Action: What Have We Learned from 13,676 Offenders and 97 Correctional Programs?" Crime and Delinquency 52(1): 77-93.

Lowenkamp and Latessa (2004). "Understanding the Risk Principle: How and Why Correctional Interventions Can Harm Low-Risk Offenders." National Institute of Corrections. Available at http://nicic.org/Library/period266.

National Institute of Corrections (2001). "Thinking for a Change: Integrated Cognitive Behavior Change Program [Lesson Plans]." Available at http://nicic.org/Library/016672.

North Carolina Department of Correction (2004). "Cognitive Behavioral Intervention Case Study." Available at http://www.doc.state.nc.us/rap/CBI-Case-Study-Final-Report.pdf.

Petersilia (2005). "Strategic Reform: Implementing Evidence-Based Programs for California Offenders." Center for Evidence-Based Corrections, UC Irvine. Available at http://ucicorrections.seweb.uci.edu/pdf/present_2005_07_12_xerox.pdf.

Rollnick and Miller (1995). "What is Motivational Interviewing?" *Behavioural and Cognitive Psychotherapy*, 23, 325-334. Available at http://motivationalinterview.org/clinical/whatismi.html.

Topics in Community Corrections, Annual Issue (1999). Classification and Risk Assessment. Available at http://nicic.org/Library/period166.
Contents include:
- Jones, Johnson, Latessa and Travis. "Case Classification in Community Corrections: Preliminary Findings from a National Survey."
- Domurad. "So You Want to Develop Your Own Risk Assessment Instrument."
- Hinzman. "The Matrix: Matching the Offender with Treatment Resources."

- Derrick, Bancroft, and Cirincione. "Navigating a System of Graduation Sanction: Adopting an Assessment Methodology for Community Corrections in New York State."
- Just. "Management Criteria for a Risk Classification Instrument in Community Corrections: A Juvenile Focus."

Experimental Design

Print references

Asher (1983). Causal Modeling. Beverly Hills, CA: Sage.

Campbell and Stanley (1963). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Cook, T.D. and Campbell, D.T. (1979). Quasi-Experimentation: Design and Analysis for Field Settings. Rand McNally, Chicago, Illinois.

Hagan (2000). Research Methods in Criminal Justice and Criminology. Boston: Allyn & Bacon.

Heckman (1979). Sample Selection Bias as a Specification Error. Econometrica 47(1): 153-162.

Heckman (1996). Randomization as an Instrumental Variable. The Review of Economics and Statistics 78(2): 336-341.

Senese (1997). Applied Research Methods in Criminal Justice. Chicago: Nelson Hall.

Spector (1981). Research Designs. Beverly Hills, CA: Sage.

Winship and Mare (1992). Models for Sample Selection Bias. Annual Review of Sociology 18: 327-350.

Winship and Morgan (1999). The Estimation of Causal Effects from Observational Data. Annual Review of Sociology 25: 659-706.

Internet references

Overview of experimental, quasi-experimental and non-experimental designs:
http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/
0,,contentMDK:20188242~menuPK:415130~pagePK:148956~piPK:216618~theSitePK:
384329,00.html

A similar reference about experimental and quasi-experimental designs (follow links):
http://www.socialresearchmethods.net/kb/design.htm

Another one about experimental and quasi-experimental designs:
http://www.ericdigests.org/1999-1/true.html

Urban Institute's straightforward description of experimental designs:
http://www.urban.org/toolkit/data-methods/experimental.cfm

Urban Institute's straightforward overview of quasi-experimental designs:
http://www.urban.org/toolkit/data-methods/quasi-experimental.cfm

Brief technical overview of quasi-experimental designs:
http://www.csulb.edu/~msaintg/ppa696/696quasi.htm

Sampling terminology explained:
http://www.socialresearchmethods.net/kb/sampterm.htm

The "sampling" section of an online statistics glossary:
http://www.cas.lancs.ac.uk/glossary_v1.1/samp.html

A Powerpoint presentation about propensity score methods:
http://sswnt5.sowo.unc.edu/VRC/Lectures/PSM_SSWR_2004.pdf

Urban Institute's coverage of propensity score methods:
http://www.urban.org/toolkit/data-methods/propensity.cfm

Brief description of the Heckman two-step procedure:
http://www11.hrdc-drhc.gc.ca/pls/edd/QEE_78016.htm

A more detailed description of Heckman:
http://home.planet.nl/~smits.jeroen/selbias/Heckman-SPSS.doc

Regression discontinuity references (straightforward descriptions):
http://www.socialresearchmethods.net/kb/quasird.htm
http://www.socialresearchmethods.net/kb/statrd.htm

Urban Institute on regression-discontinuity:
http://www.urban.org/toolkit/data-methods/regression.cfm