DRUG COURT REVIEW

Volume V, Issue 2

Special Research Edition

NATIONAL DRUG COURT INSTITUTE
ALEXANDRIA, VIRGINIA

DRUG COURT REVIEW

Volume V, Issue 2
NATIONAL DRUG COURT INSTITUTE

Judge Karen Freeman-Wilson (Ret.), Executive Director
C. West Huddleston, III, Director
4900 Seminary Road, Suite 320
Alexandria, Virginia 22311
Tel. (703) 575-9400
Fax (703) 575-9402
www.ndci.org

NDCI is supported by the Office of National Drug Control Policy, Executive Office of the President and the Bureau of Justice Assistance, U.S. Department of Justice.

This document was prepared under Cooperative Agreement Number 2005-DC-BX-K003 from the Bureau of Justice Assistance, U.S. Department of Justice, with the support of the Office of National Drug Control Policy, Executive Office of the President. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice or the Executive Office of the President.

Printed in the United States of America.

Drug courts perform their duties without manifestation, by word or conduct, of bias or prejudice, including, but not limited to, bias or prejudice based upon race, gender, national origin, disability, age, sexual orientation, language, or socioeconomic status.

INTRODUCTION

The Editorial Board is pleased to present the second issue of volume five of the *Drug Court Review* (Volume V, 2). This issue of Volume V is a special research edition, closely examining such pertinent issues as *performance measurement*, *process evaluation*, and *recidivism analysis* under the broad structure of a *national research agenda.*

Continued refinement of drug court research, both through program evaluation and exploration of the drug court process, is critical for the advancement of the field.

In this issue:

- Douglas B. Marlowe, J.D., Ph.D., Cary Heck, Ph.D., C. West Huddleston, III, and Rachel Casebolt set a national agenda for drug court research. Recommended by NDCI's National Research Advisory Committee and field-tested by a national sample of drug court practitioners and administrators, this strategy aims to focus future drug court research by presenting the research topics, questions, and priorities required to create a robust literature for the drug court field.
- Cary Heck, Ph.D., and Meridith H. Thanner, Ph.D., present a carefully enumerated method of drug court research known as performance measurement. The authors distill important elements of drug court research into four major categories and recommend the best research-supported methods for measuring and reporting the data for maximum impact.
- Cary Heck, Ph.D., and Meridith H. Thanner, Ph.D., present specific recommendations on the purposes, techniques, methodologies, and procedures for conducting process evaluations. Covering all the important areas ubiquitous to drug court, the authors also suggest going

beyond the basic model and providing useful idiosyncratic results for the needs of individual courts.

- Finally, Michael Rempel explores the nuances of conducting a valid recidivism analysis. Presenting complicated research and statistical concepts in plain language, this article clearly defines recidivism and its permutations; explains the universe of clients that should be considered; lists and describes the various options for comparison groups and their ramifications; and identifies methods of ensuring the validity of the comparison.

THE DRUG COURT REVIEW

Published semi-annually, the *Review*'s goal is to keep the drug court practitioner abreast of important new developments in the drug court field. Drug courts demand a great deal of time and energy of the practitioner. There is little opportunity to read lengthy evaluations or keep up with important research in the field. Yet, the ability to marshal scientific and research information and "argue the facts" can be critical to a program's success and ultimate survival.

The *Review* builds a bridge between law, science, and clinical communities, providing a common tool to all. A headnote and subject indexing system allows access to evaluation outcomes, scientific analysis, and research on drug court related areas. Scientific jargon and legalese are interpreted for the practitioner into common language.

Although the *Review*'s emphasis is on scholarship and scientific research, it also provides commentary from experts in the drug court and related fields on important issues to drug court practitioners.

THE NATIONAL DRUG COURT INSTITUTE

The *Drug Court Review* is a project of the National Drug Court Institute. NDCI was established under the auspices of the National Association of Drug Court Professionals and with the support of the Office of National Drug Control Policy, Executive Office of the President, and the Bureau of Justice Assistance, U.S. Department of Justice.

The National Drug Court Institute's mission is to promote education, research, and scholarship to the drug court field and other court-based intervention programs.

Historically, education and training in the drug court field have only been available at regional workshops and the annual national conference; analysis and scholarship were largely limited to anecdotes and personal accounts.

That situation has changed. Evaluations exist on dozens of drug court programs. Scholars and researchers have begun to apply the rigors of scientific review and analysis to the drug court model. The level of experience and expertise necessary to support such an institution now exist.

Since its creation in December 1997, NDCI has launched a comprehensive practitioner training series for judges, prosecutors, public defenders, court coordinators, treatment providers, and community supervision officers; developed a research division responsible for developing a scientific research agenda and publication dissemination strategy for the field, as well as developing a series of evaluation workshops; and published a monograph series on relevant issues to drug court institutionalization and expansion.

ACKNOWLEDGEMENTS

I wish to thank all those who have contributed to this issue of the *Drug Court Review*: to the Office of National Drug Control Policy, Executive Office of the President, and the Bureau of Justice Assistance, U.S. Department of Justice, for the leadership, support, and collaboration that those agencies have offered to the National Drug Court Institute; and to Dr. Douglas B. Marlowe, Dr. Cary Heck, C. West Huddleston, III, Rachel Casebolt, Dr. Meridith H. Thanner, and Michael Rempel for their contributions as authors.

Judge Karen Freeman-Wilson (Ret.)
Executive Director
National Drug Court Institute

CONTENTS

A NATIONAL RESEARCH AGENDA FOR DRUG COURTS: PLOTTING THE COURSE FOR SECOND-GENERATION SCIENTIFIC INQUIRY

By Douglas B. Marlowe, J.D., Ph.D.,
Cary Heck, Ph.D.,
C. West Huddleston, III, and
Rachel Casebolt
National Drug Court Institute,
Alexandria, VA

In 2005, the National Drug Court Institute convened an expert panel of nationally recognized scholars to develop a research agenda for adult drug courts. Named the National Research Advisory Committee (NRAC), this expert panel specified standardized criteria for identifying critical research questions for the field, which ultimately led to the development of a list of 23 research priorities for drug court. Subsequently, this list of priorities underwent field review by a national sample of drug court practitioners and administrators, who endorsed each of the research topics and rank-ordered them in importance. This research agenda reflects the considered opinion of both scholars and practitioners in the drug court field about the important research topics that need to be addressed, sets priorities for researchers and evaluators about which questions to focus on, and provides a road map for funders and sponsors for identifying those research proposals that are most relevant to drug court practices and policies. Future NRAC meetings are planned that will focus on juvenile drug courts, DUI/DWI courts, and family dependency treatment courts.

Douglas B. Marlowe, J.D., Ph.D., is the Director of Law & Ethics Research at the Treatment Research Institute (TRI), and an Adjunct Associate Professor of Psychiatry at the University of Pennsylvania School of Medicine. His research focuses on examining the role of coercion in drug abuse treatment, the effects of drug courts and other

diversion programs for drug abusing offenders, and behavioral treatments for drug abusers and offenders.

Cary Heck, Ph.D., is an Assistant Professor of Criminal Justice at the University of Wyoming as well as the Director of Research for the National Drug Court Institute. Previously, he served as Statewide Director of Drug Courts for Louisiana through the State Supreme Court. Dr. Heck has also worked as a small town police officer, probation officer, gang task force member, and counselor for delinquent youth.

C. West Huddleston, III is the Director of the National Drug Court Institute. He is a Board-licensed counselor with 13 years of clinical experience at the county, state, and federal levels. Previously, Mr. Huddleston worked throughout the Tennessee and Oklahoma correctional systems developing and managing substance abuse units, as well as co-developing and administering two drug courts.

Rachel Casebolt is the Research Coordinator at the National Drug Court Institute. As part of her duties, she works with federal implementation and enhancement grantees to ensure that guidelines for local drug court research are observed.

Direct all correspondence to Douglas B. Marlowe, J.D., Ph.D., Treatment Research Institute at the University of Pennsylvania, 600 Public Ledger Bldg., 150 S. Independence Mall West, Philadelphia, PA 19106-3475. (215) 399-0980; (215) 399-0987 (fax); **Marlowe@tresearch.org**.

ARTICLE SUMMARIES

PAST THE FIRST GENERATION OF RESEARCH

[1] The first generation of drug court research focused on the question of whether drug courts work. Having affirmatively answered that question, future research should concentrate on the mechanisms behind their success.

NATIONAL RESEARCH ADVISORY COMMITTEE

[2] Developed to guide future drug court research nationally, the committee defined selection criteria through which to guide research priorities and identified key content areas.

NATIONAL RESEARCH AGENDA

[3] The committee focused its recommendations on specific questions of long term outcomes, impacts on minorities, judicial practices, incentives and sanctions, treatment services, case management, and issues of collaboration.

CONCLUSION

[4] While drug courts have achieved initial success, their ultimate success depends on their ability to self-examine and refine the model.

INTRODUCTION

Drug courts reduce criminal recidivism by roughly 15 to 20 percentage-points as compared to the traditional adjudication of drug-related offenses, and they enhance offenders' exposure to substance abuse treatment nearly six-fold as compared to standard or intensive probationary conditions. These are the consistent conclusions reached by numerous research scholars (e.g., Belenko, 1998; Cissner & Rempel, 2005; Goldkamp, 2003; Harrell, 2003; Marlowe, DeMatteo, & Festinger, 2003; Roman, Townsend, & Bhati, 2003) and endorsed by the U.S. Government Accountability Office (GAO, 2005) and the White House Office of National Drug Control Policy (ONDCP, 2005). Importantly, these robust effects have been sustained under stringent experimental research conditions (e.g., Gottfredson, Najaka, & Kearley, 2003). Moreover, it appears that the magnitude of the effects increase even further when drug court services are appropriately targeted to the needs of the most incorrigible, high-risk drug offenders (e.g., Fielding, Tye, Ogawa, Imam, & Long, 2002; Marlowe, Festinger, Lee, Dugosh, & Benasutti, 2006; Rempel & DeStefano, 2001).

[1] On the heels of these first-generation research findings, drug courts have grown at an exponential rate from single-digit numbers in the early 1990s to over 1,600 drug courts by the end of 2004 (Huddleston, Freeman-Wilson, Marlowe, & Roussell, 2005). The concern at this juncture is that growth could be outpacing data. More information is needed to determine how to tailor program resources most efficiently, how to target inclusion criteria to the most suitable candidates, and how to modify drug court interventions to elicit more robust and longer-term effects. Scholars are now calling out in unison for a second generation of research focusing no longer on *whether* drug courts work, but rather on *how* and *for whom* they work, and how they might work even better (e.g., Cissner & Rempel,

2005; Marlowe et al., 2003; Goldkamp, 2001; Longshore et al., 2001).

Several research portfolios do exist that are investigating issues of relevance to drug court practice. For instance, studies of behavioral and pharmacological interventions for addiction are yielding critically important findings that promise to improve outcomes for drug court participants. Other studies investigating alternative sentencing strategies for drug offenders promise to identify less costly and less punitive dispositions that can still serve public-safety and public-health objectives.

But this is not enough. Few studies are addressing the specific mechanism(s) of action believed to be responsible for drug courts' superior effects. Combining ongoing judicial supervision with evidence-based clinical services, intensive case management, community-based corrections, and operant conditioning techniques has leveraged outcomes not heretofore seen with this intransigent population. If drug courts exceed the sum of their parts by integrating treatment and correctional interventions in synergistic ways, the existing body of research evidence fails to do justice to this new paradigm and pales in its efforts to advance the field. Studying outmoded models of "treatment vs. punishment" cannot be expected to shed light on this new way of doing business for the courts, which rejects the exclusive embrace of those single-minded strategies as too simplistic to solve the complicated problem of drug-related crime. Studies are needed to investigate the defining ingredients of drug court programs and to clearly distinguish the drug court model from past endeavors that have produced lackluster results.

NATIONAL RESEARCH ADVISORY COMMITTEE

[2] Recognizing the critical need to conduct research aimed at these nuanced issues, the National Drug Court

Institute (NDCI) convened a nationally recognized expert panel of drug court researchers and practitioners called the National Research Advisory Committee (NRAC). The committee met repeatedly throughout 2005 to develop procedures for identifying essential second-generation research questions, which ultimately resulted in a national drug court research agenda (see Table 1, Appendix). Under the direction of Dr. Douglas Marlowe from the Treatment Research Institute (TRI) at the University of Pennsylvania and Dr. Cary Heck from the University of Wyoming, NRAC developed an objective means for specifying research priorities, raising the most pressing legal and social issues facing drug courts.

Selection Criteria. NRAC first identified standardized criteria to guide the specification of research priorities, as listed below. These criteria were not rank-ordered in importance and no one research question was expected to satisfy every criterion. The goal was to consider potential research questions in light of each of these important criteria.

- Is the research question policy-relevant? For example, does it have implications for such matters as sentencing or dispositional policies?
- Is the question of substantial interest to practitioners and relevant to their day-to-day activities?
- Does it address one or more of the Ten Key Components of problem solving courts (NADCP, 1997)?
- Is the question amenable to high-quality, controlled research? For example, is the hypothesis falsifiable and could random assignment be feasible?
- Is the research likely to have a “high payoff” potential for the field as a whole? For example, is it apt to lead to best-

practice recommendations or to professional credentialing standards?

- Does it build logically upon an existing body of research evidence that would suggest it is likely to bear fruit in terms of positive findings?
- Can a potential funding source or sponsor for the research be identified? Is the question already the subject of substantially funded research?
- Is the question innovative in terms of involving new clinical applications or new research methodologies?
- Does the matter relate to the integration or synergy between public safety (criminal justice) and public health (treatment) perspectives, or does it simply focus on one of these perspectives within the context of a drug court?

Content Areas. NRAC further decided that the research priorities should address each of the broad content areas represented within drug courts and other problem solving court programs. Because drug courts reflect a unique blending of various systems and approaches, and are committed to providing equal access to the courts for all citizens, it was felt that key issues should be investigated within each of the following domains:

- Judicial or court practices
- Incentives and sanctions
- Substance abuse treatment and other services
- Community supervision and case management
- Inter-agency and inter-system collaboration
- Differential impacts on minority citizens

Field Review. Based upon the above criteria, NRAC identified a list of 23 research priorities for drug courts (see

Table 1, Appendix). Subsequently, this list underwent field review by a national sample of drug court administrators and practitioners. NDCI maintains a list of primary points of contact (PPOCs) in every state and territory in the U.S., consisting of presidents of state drug court associations, statewide drug court coordinators, or individuals with ongoing responsibility for administering drug court programming within a particular jurisdiction's Administrative Office of the Courts (AOC). The PPOCs forwarded a Likert-scale survey derived from the NRAC recommendations (see Table 1, Appendix) to their respective Statewide Drug Court Steering Committee, State Drug Court Commission, or comparable governing body. These committees, in turn, distributed the surveys to drug court practitioners and administrators within their respective jurisdictions.

Each research priority identified by NRAC was rated by respondents in terms of its perceived importance to the field on a 5-point Likert scale, ranging from 0 ("not at all important") to 5 ("extremely important"). Additional space was provided for respondents to identify other research priorities. NDCI followed up with each PPOC via telephone or e-mail to ensure that every state and territory was represented in the survey. Descriptive data on the survey results are presented in Table 1 in the Appendix.

The respondents (N = 150) reflected a broad range of professional disciplines represented within drug court programs in every jurisdiction, including drug court coordinators, judges or magistrates, prosecutors, defense attorneys, treatment providers, case managers, community corrections officers, and court clerks or administrators.

To this point, NRAC has focused on adult criminal drug courts, because these are the most prevalent type of problem solving court program, have been in existence the longest, and currently have the largest body of research evidence supporting their efficacy. In subsequent meetings,

the committee has been engaged in identifying research priorities for other prevalent types of problem solving courts, including DUI/DWI courts, juvenile drug courts, and family dependency treatment courts and is committed to presenting its findings and recommendations to the field.

A NATIONAL RESEARCH AGENDA FOR DRUG COURTS

Long Term Outcomes

[3] *Research Question 1*: What are the long-term effects of drug courts and other problem solving courts on other important outcomes, such as substance use, psychological health, physical health, employment, or parenting? What components of the drug court model contribute to the most effective outcomes in those areas?

Most of the existing research on drug courts has relied on official arrest and conviction records to measure outcomes. This is because evaluators can usually gain ready access to state criminal justice databases at a manageable cost. Unfortunately, relatively little is known about the effects of drug courts on other client-level outcomes, such as substance use, family interactions, employment, and medical or psychiatric functioning. This is because it is often difficult and expensive to track down participants for purposes of administering post-treatment interviews or urine drug tests. Yet it is important to recognize that the basic logic-model of drug courts assumes that substance abuse often mediates criminal activity; therefore, treating addiction is believed to elicit sustained reductions in criminal recidivism. Without measuring effects on substance use and other psychosocial indicators, it is not possible to test this central hypothesis of drug courts. Research is required that measures the effects of drug courts on a wider range of client-level variables, and that permits researchers an understanding of how drug courts exert effects on both mediating and distal outcomes.

Differential Impacts on Minority Citizens

Research Question 2: Do minority sub-groups have differential access to drug-court programs or differential success or failure rates? Are they subjected to different types or amounts of sanctions or rewards for comparable performance? Do they receive different types of treatment services? If so, why and how can drug courts correct this?

Racial, ethnic, and cultural minorities are disproportionately represented in the criminal justice system. Minorities generally do not report higher rates of illicit drug use in anonymous national household surveys, yet African Americans are imprisoned at nearly four times the rate of Caucasians for drug-related offenses and Hispanics are imprisoned at more than twice the rate of Caucasians (e.g., Iguchi, Bell, Ramchand, & Fain, 2005). Concerns have been raised about whether citizens of color have equivalent access to resource-intensive drug court programs, and whether the services offered within drug courts are culturally responsive and account for the unique obstacles often faced by people of color, not the least of which may include language barriers and culturally divergent concepts of wellness or healing. Research is needed to address issues related to access and performance of minority groups within drug court programs.

Judicial and Court Practices

Research Question 3: How are outcomes influenced by having a permanently dedicated drug court judge and docket, as opposed to annually rotating assignments?

In many jurisdictions, drug courts were founded by innovative and committed judges who continued to lead the program for years. In other jurisdictions, judges may be assigned to drug court on an annually rotating docket or based upon their seniority. It is important to determine whether such arrangements influence the effectiveness of

drug court programs. This information will provide needed guidance to President Judges about how to most effectively mete out judicial assignments for drug court dockets.

Research Question 4: Does it matter whether a judge wants to be in drug court, or can any judge be "brought up to speed"?

It is commonly believed that the success of drug courts hinges, at least in part, on the dedication of its staff members—particularly on that of the judge who leads the team and sets the tone for the program during status hearings. This raises concerns about whether regular criminal court judges who may have little interest in drug court can be brought in on short notice to provide continuing oversight to a program, or whether a ready cadre of dedicated personnel is necessary to maintain continuity. Research is needed to determine how the motivation and training of judges influence client outcomes.

Research Question 5: What traits or characteristics of the judge, if any, are associated with better outcomes for various clienteles?

Similar to questions that have been raised about the therapeutic alliance in psychotherapy, questions have been raised about how drug court judges should interact with clients during status hearings, and how these interactions might be influenced by a judge's personality or relational style. Research is needed to determine what attributes of a judge make him or her better suited for drug court practice, and whether this might vary by the nature of the client population.

Research Question 6: Can equivalent outcomes be attained using alternative judicial arbiters, such as masters or commissioners? Does this vary by clientele?

Judges cost money. Moreover, in many jurisdictions, judges are elected to the bench by the public or are appointed by the executive branch of government with the advice and consent of the legislature. As a result, judges are frequently in short supply, and they may be hesitant to make unpopular treatment-oriented decisions that can be seen as "soft on crime." Many states also have lower tiers of judicial arbiters called "magistrates" or "commissioners," who may sit on the bench of a limited-jurisdiction court or render a restricted range of decisions. For example, a magistrate might be empowered to oversee judicial status hearings in drug court and make limited decisions related to treatment and supervisory conditions; however, they may be required to bring in a judge to rule on more serious matters such as termination from drug court or imposition of jail sanctions. Research is needed to determine whether such arrangements can be effective and cost-efficient, and how they might influence such issues as clients' perceptions of procedural fairness or due process.

Incentives and Sanctions

Research Question 7: What are the impacts of brief jail sanctions ("flash incarceration") on clients who are noncompliant with their care plans or program requirements? Do these impacts vary by the nature of the clientele (e.g., adults vs. juveniles, or criminal offenders vs. parents in dependency cases)? Do they vary by the target behavior (e.g., non-attendance in counseling vs. relapse to drug use)?

Drug court judges have substantial authority to impose potent sanctions and rewards contingent upon offenders' conduct in treatment. Most notably, they may be authorized to impose brief intervals of jail detention (known as "flash incarceration") for noncompliance in treatment or unremitting substance use. Not surprisingly, jail sanctions are among the most controversial aspects of judicial intervention in drug courts. For instance, major legislative

policy debates are underway in California with regard to proposed amendments to "Proposition 36," which would authorize brief jail sanctions for noncompliant probationers who had been diverted into treatment in lieu of incarceration. The Supreme Court of California is considering a similar issue in a family dependency treatment court case (*In Re Olivia J.*), which involves the propriety of using civil contempt powers to briefly jail non-compliant parents in child neglect proceedings. It is essential to provide controlled research data regarding the effects and potential side-effects of brief jail sanctions in drug courts and other problem solving court programs.

Research Question 8: Are there sub-groups of drug court clients for which "rational" models of rewards and sanctions are differentially effective or need to be substantially modified, such as mentally ill offenders, juveniles, or psychopaths?

Research indicates that certain high-risk drug offenders may respond differently to sanctions and rewards than other individuals. For example, youthful offenders and those with antisocial personality disorder tend to discount the probability of receiving a serious sanction in the long-term in favor of earning an immediate reward (e.g., Patterson & Newman, 1993). They are also more likely to opt for smaller short-term rewards on delay-discounting tasks than to forestall gratification in favor of larger rewards to be earned in the future (e.g., Petry, 2002). This apparent hypersensitivity to rewards, imperviousness to sanctions, and impulsivity could reflect executive-control deficits stemming from damage or immaturity to the pre-frontal cortex (e.g., Fishbein, 2000). It has also been observed, anecdotally, that mentally ill offenders may react in unanticipated ways to negative sanctions. Research is needed to determine whether drug courts should substantially modify their slates of sanctions and rewards or apply them differently for certain subgroups of drug offenders.

Research Question 9: How are outcomes affected when a drug court imposes a pre-defined "matrix" of sanctions or rewards, as opposed to individualizing its responses and keeping clients "guessing"?

Many drug courts develop a pre-specified slate or matrix of graduated sanctions and rewards that are applied in response to successive infractions or accomplishments in the program. This matrix may be listed in a program manual to give clients clear advance warning about the types of consequences that can be imposed, and to enhance clients' sense of procedural justice and fairness in the program. Other drug courts prefer to craft their responses on an individualized basis, in consideration of each client's unique clinical and criminogenic needs. Unfortunately, research provides little basis for determining which approach may be most effective and for which types of clients.

Substance Abuse Treatment and Other Services

Research Question 10: What is the optimum length of time for required participation in a drug court program? Does this vary by clientele or by the drug of choice? Does drug court "accelerate" recovery because of the additional services and monitoring?

Most drug courts are scheduled to be a minimum of 12 to 18 months in duration and require clients to satisfy fairly stringent criteria for graduation. It is an open question whether such a standardized course of treatment, with standardized completion criteria, is a suitable approach for all clients and whether a year-long regimen might be excessively burdensome or costly in some cases. It is also an open question whether the intensive and multi-faceted services offered in drug courts might elicit faster gains than are typically seen in traditional community-based substance abuse treatment programs. Research is needed to identify the ideal term(s) for drug court programs, and to determine

whether drug courts may speed up the recovery process as compared to traditional modalities of substance abuse treatment.

Research Question 11: What are the most effective continuing-care strategies that result in the greatest likelihood of long term success, focusing specifically on practices that (a) utilize the continued influence of the criminal justice system following completion of the drug court program and (b) are embraced by program graduates after the drug court no longer has active jurisdiction over their case?

It is common practice for drug courts to relax clients' treatment and supervisory obligations as they near completion of the program, and the requirements often end precipitously following graduation. This could lead to a rapid decline in functioning for some individuals upon discharge. Research is needed to identify the most effective continuing care strategies that can be utilized when the drug court no longer has legal jurisdiction over the case.

Research Question 12: Are there sub-types of drug offenders who could benefit from the monitoring components of the drug-court model, even if they are not actively addicted to drugs or alcohol and may not require formal substance abuse treatment?

The drug court model assumes that most drug offenders are addicts or serious drug abusers and that drug use fuels or exacerbates other criminal activity. As a result, drug court clients are typically required to satisfy an intensive regimen of treatment and supervisory obligations. However, research suggests that roughly one-third of clients in drug courts do not have a diagnosable or clinically significant substance use disorder (e.g., DeMatteo, Marlowe, & Festinger, 2006). For these clients, standard drug court services may be ineffective or unduly costly. Instead, these low-risk clients may be best suited for a secondary prevention

approach directed at interrupting the acquisition of addictive behaviors. Alternatively, they might respond to a "coerced abstinence" model that simply focuses on holding them accountable for drug-positive urine samples (e.g., Kleiman et al., 2003). Research is needed to identify those clients who may respond to the monitoring elements of drug courts without the necessity of providing an entire menu of costly and intrusive clinical services.

Research Question 13: What additional or adjunctive services are most related to positive outcomes in drug courts and most likely to serve public-safety aims? In particular, should employment or educational attainments be required prior to graduation from drug courts?

Drug courts were created to reduce drug use and crime. It is uncertain whether drug courts must, or should, intervene further against the myriad other problems clients frequently present with in order to maintain treatment gains. For example, if unemployment or family dysfunction is apt to precipitate relapse or recidivism, then drug courts might not be able to accomplish their primary task unless they also improve these other problems as well. On the other hand, expecting too much from clients or overburdening them with an array of services could undermine treatment goals. Research is needed to determine the circumstances under which adjunctive services improve drug- and crime-related outcomes at a manageable cost to the program.

Research Question 14: Are there differential effects when a drug court requires an abstinence-only policy from the outset, as opposed to following a "harm reduction" approach that approximates abstinence over time?

Some drug courts view illicit drug use as unacceptable behavior that cannot be tolerated by law enforcement authorities and presents an unwarranted risk to public safety. These courts may also believe that such

voluntary misconduct can be readily brought under behavioral control through the stringent application of rewards and sanctions. Other drug courts view addiction as a disease with compulsive features that takes some time to treat, and that will inevitably be characterized by relapse. Research is needed to compare outcomes across programs holding divergent views about the effects of punishing substance use, and more importantly, to determine which types of clients respond better to different enforcement procedures.

Community Supervision and Case Management

Research Question 15: What types of clients require frequent judicial contacts, and what types can be effectively and safely managed by community corrections officers, probation or parole officers, case managers, or treatment providers?

Evidence suggests that certain types of low-risk drug offenders can be supervised safely and effectively using community corrections officers or treatment providers, whereas high-risk offenders require the authority and power of a judge to bring their substance use and illicit activity under control (Marlowe et al., 2006). Further research is needed to evaluate the generalizability of these findings for the full range of drug-possession offenders. This will permit communities to preserve their precious judicial resources while safeguarding public safety and contributing to better outcomes for their clients.

Research Question 16: What types of community monitoring technologies (e.g., anklet monitors, Secure Continuous Remote Alcohol Monitor (SCRAM), phone monitors, patches) and practices (e.g., surprise home visits) are associated with better outcomes in drug courts? Do these outcomes vary by population? How do these technologies and practices affect clients' perceptions of such things as "procedural justice" or "perceived deterrence"?

Drug courts involve a close partnership between the courts, community corrections officers, and treatment providers. Research is needed to determine which types of community-based monitoring practices can be most effectively managed by probation and parole officers, and which can be most effectively integrated with judicial practices and standard clinical interventions.

Inter-Agency and Inter-System Collaboration

Research Question 17: How can we develop better methods and instrumentation to measure the degree of collaboration between agencies and systems in drug courts?

As discussed earlier, the effects of drug courts appear to exceed the sum of their parts by creating a synergy between clinical, judicial, and correctional interventions. Unfortunately, few instruments exist to measure the degree to which this integration has been achieved in a particular program, and to pinpoint the nature of effective cross-agency interactions. The drug court field needs better assessment tools to measure these synergistic and collaborative processes.

Research Question 18: Are outcomes affected by having clinicians and case managers appear during status hearings to give testimony, as opposed to sending written reports or transmitting data elements? Does this affect clients' perceptions of the therapeutic alliance and their willingness to disclose important personal information?

The time that it takes for clinicians to appear at status hearings in court is time taken away from other important functions, such as providing treatment services. On the other hand, having clinicians appear in court can serve to plug gaps in communication and ensure that sanctions and rewards are applied by the judge with the requisite certainty and immediacy that is necessary for effective outcomes.

Research is needed to determine the circumstances under which clinicians' presence in court is most likely to improve outcomes while having the fewest negative effects on the therapeutic alliance.

Research Question 19: Are outcomes affected by having clinicians share only limited data elements (e.g., counseling attendance and drug-testing results) with the court and other professionals, as opposed to sharing a wider range of clinical information? Does this affect clients' perceptions of the therapeutic alliance and their willingness to disclose important personal information?

The intrusion of a judge into the therapeutic relationship could be disruptive or harmful under some circumstances. Clients may be hesitant, for example, to confide clinically important information to their therapists for fear the information will be disclosed to the judge and used against their legal interests. On the other hand, having clinicians provide detailed progress reports to the judge prevents clients from "falling through the cracks" and eluding deserved sanctions or losing deserved rewards. Research is needed to determine the appropriate scope of information-sharing that permits effective communication among drug court staff members, while at the same time preserving the sanctity and trust of the therapeutic relationship.

Research Question 20: Are decisions more consistent and outcomes more effective when the judge acts as the final arbiter of clients' performance during status hearings, or when the team reaches a general consensus on such matters?

Drug courts are designed to operate on a team basis, with the judge conceptualized as a leader among equals (NADCP, 1997). Constitutional due process requires the judge to exercise final and independent judgment on all matters influencing a client's legal status and rights; however, it is appropriate for the judge to rely on the expertise of other

professionals in making these decisions, and the judge must at least consider the arguments of legal counsel on both sides of the case. Unfortunately, research is virtually nonexistent for understanding how judges can make the most informed decisions. For example, studies are needed to determine whether reliance on a team consensus leads to better client outcomes or more correct decisions, or whether independent judgment is ultimately more reliable.

Research Question 21: Does participation in drug court raise the quality of all staff members' performance, such as improving the quality of treatment?

Drug court judges have substantial prestige and influence within their communities, and it is hoped that by partnering with clinicians and correctional professionals, they will enhance the performance of all parties involved in the drug court process. Research is needed to determine how judges can be the most effective "consumers" or "purchasers" of substance abuse treatment services and probation or parole services.

Research Question 22: Are outcomes improved or are services more efficient when the drug court coordinator is an agent of the court system, the treatment system, probation/parole, or some combination of these?

Virtually all drug court programs have a designated drug court coordinator who is primarily responsible for coordinating the services across state agencies, providing for staff training and supervision, and purchasing materials and services for the program. In some states, this individual might be an agent of the substance abuse treatment system, whereas in other jurisdictions he or she might be an employee of the AOC or of community corrections. Research is needed to determine whether client outcomes, inter-agency collaboration, or even the political influence of a program within a jurisdiction are enhanced when the drug court

coordinator is an agent of a particular state agency or has a specific professional identity or level of professional training.

Research Question 23: What methods of client staffing or case conferences lead to the most effective sharing of information and to the best outcomes? In particular, is it necessary for the prosecution and defense to be present during staffing and during court proceedings?

Prior to holding status hearings in court, drug court team members typically meet in a case conference or case review to share information about clients and reach decisions about how to respond to clients' performance in the program. In some drug courts, these case reviews focus on dealing with the most difficult cases, whereas others might focus on staging therapeutic interactions for open court, or generally improving team communication and processes. Some programs solicit regular input from all drug court staff members during these meetings, including the defense and prosecution, whereas others may use the time for the judge and clerk to process paperwork or plan the court calendar, or for treatment providers to report on clinical information about their cases. Research is needed to determine the most effective and efficient ways to hold case reviews and to reach team decisions on important matters for the program.

CONCLUSION

There are two kinds of success: initial and ultimate.
—Winston Churchill

[4] Drug courts have achieved undeniable initial success. In the span of only 15 years, they have evolved from an anomalous experiment within a few courts to a major movement within the criminal justice system. There are now more than 1,600 drug courts and over 2,500 total problem solving courts located in every jurisdiction in the U.S. (Huddleston et al., 2005) as well as several foreign nations.

More research has now been published on the positive effects of drug courts than on virtually all other interventions for drug-abusing offenders combined, including reviews of more than 100 program evaluations (Belenko, 1998; 2001), randomized experimental studies (Gottfredson et al., 2003; Turner, Greenwood, Fain, & Deschenes, 1999), statewide systems evaluations (e.g., Rempel et al., 2003), and national recidivism estimates (Roman et al., 2003). No other criminal justice program can come close to boasting this level of programmatic success or scientific productivity.

Yet, the future is far from secure. Cuts in state and federal funding threaten the integrity of existing programs and are slowing down the development and expansion of new slots. Competing models, such as California's Proposition 36, claim that success can be achieved by eliminating many of the defining attributes of drug courts, not the least of which include judicial monitoring and graduated sanctions and incentives. In addition, because drug courts tend to be resource-intensive, it is difficult for them to serve a wide-ranging proportion of the drug-involved offender population. Guidance is critically needed to determine which segments of the drug-offender population are best suited for drug courts, and to indicate how certain aspects of the drug court model might be infused into the practices and philosophy of general-jurisdiction criminal courts.

It is too late in the day to rehash old arguments about whether drug courts work. Critics who ignore the current cache of evidence supporting drug courts are unlikely to be swayed by more of the same data. The time has come to move the field forward to a new generation of more sophisticated research questions: Which types of offenders are bested suited to drug court, what types of services within drug court contributes to the most effective outcomes, what is the mechanism of action that explains the superior effects of drug court, and how can certain principles and practices of drug court be extended to the larger criminal justice context?

Answers to these questions will point the field in important directions towards improving clinical practice, drug policy, and public safety.

Although drug courts have achieved initial success, their ultimate success depends on their ability to answer these nuanced and sophisticated research questions—and to do so in an atmosphere of competing, and sometimes mutually inconsistent, research priorities. Research portfolios that take the single-minded perspective that drug abuse is simply a disease requiring treatment are unlikely to pursue the critical avenues of research necessary to unlock the synergistic ingredients of drug court programs. Similarly, those that view drug abuse as simply unlawful conduct are unlikely to add new knowledge to the field. New paradigms call for new research methods and new research questions. Only then will the ultimate success of drug courts be secured.

REFERENCES

Belenko, S. (1998). Research on drug courts: A critical review. *National Drug Court Institute Review 1*(1), 1-42.

Belenko, S. (2001). *Research on drug courts: A critical review: 2001 update*. New York: National Center on Addiction and Substance Abuse at Columbia University.

Cissner, A. B., & Rempel, M. (2005). *The state of drug court research: Moving beyond "do they work?"* New York: Center for Court Innovation.

DeMatteo, D.S., Marlowe, D.B., & Festinger, D.S. (2006). Secondary prevention services for clients who are low risk in drug court: A conceptual model. *Crime & Delinquency 52*, 114-134.

Fielding, J.E., Tye, G., Ogawa, P.L., Imam, I.J., & Long, A.M. (2002). Los Angeles County drug court programs: Initial results. *Journal of Substance Abuse Treatment 23*, 217-224.

Fishbein, D. (2000). Neuropsychological function, drug abuse, and violence: A conceptual framework. *Criminal Justice & Behavior 27*, 139-159.

Goldkamp, J. (2001). Do drug courts work? Getting inside the drug court black box. *Journal of Drug Issues 31*, 27-72.

Goldkamp, J. (2003). The impact of drug courts. *Criminology & Public Policy 2*, 197-206.

Gottfredson, D.C., Najaka, S.S., & Kearley, B. (2003). Effectiveness of drug treatment courts: Evidence from a randomized trial. *Criminology & Public Policy 2*, 171-196.

Harrell, A. (2003). Judging drug courts: Balancing the evidence. *Criminology & Public Policy 2*, 207-212.

Huddleston, C.W., Freeman-Wilson, K., Marlowe, D.B., & Roussell, A.P. (2005, May). *Painting the Current Picture: A National Report Card on Drug Courts and Other Problem Solving Courts, I*(2). Alexandria, VA: National Drug Court Institute, National Association of Drug Court Professionals.

Iguchi, M.Y., Bell., J., Ramchand, R.N., & Fain, T. (2005). How criminal system racial disparities may translate into health disparities. *Journal of Health Care for the Poor and Underserved 16*, 48-56.

In Re Olivia J., 124 Cal. App. 698, 21 Cal. Rptr. 3d 506 (Cal. Ct. App. 2004), *cert. granted*, 108 P.3d 862 (Cal. March 16, 2005) (S130457).

Kleiman, M.A.R., Tran, T.H., Fishbein, P., Magula, M.T., Allen, W., & Lacy, G. (2003). *Opportunities and barriers in probation reform: A case study of drug testing and sanctions*. Berkeley, CA: University of California, California Policy Research Center.

Longshore, D., Turner, S., Wenzel, S., Morral, A., Harrell, A., McBride, D., Deschenes, E., & Iguchi, M. (2001). Drug courts: A conceptual framework. *Journal of Drug Issues 31*, 7-25.

Marlowe, D.B., DeMatteo, D.S., & Festinger, D.S. (2003). A sober assessment of drug courts. *Federal Sentencing Reporter 16*, 153-157.

Marlowe, D.B., Festinger, D.S., Lee, P.A., Dugosh, K.L., & Benasutti, K.M. (2006). Matching judicial supervision to clients' risk status in drug court. *Crime & Delinquency 52*, 52-76.

National Association of Drug Court Professionals. (1997). *Defining drug courts: The key components*. Washington, DC: Office of Justice Programs, U.S. Department of Justice.

Patterson, C.M., & Newman, J.P. (1993). Reflectivity and learning from aversive events: Toward a psychological mechanism for the syndromes of disinhibition. *Psychological Review 100*, 716-736.

Petry, N.M. (2002). Discounting of delayed rewards in substance abusers: Relationship to antisocial personality disorder. *Psychopharmacology 162*, 425-432.

Rempel, M., & DeStefano, C.D. (2001). Predictors of engagement in court-mandated treatment: Findings at the Brooklyn Treatment Court, 1996-2000. *Journal of Offender Rehabilitation 87*, 124.

Roman, J., Townsend, W., & Bhati, A. (2003). *National estimates of drug court recidivism*. Washington, DC: National Institute of Justice, U.S. Department of Justice.

Turner, S., Greenwood, P., Fain, T., & Deschenes, E. (1999). Perceptions of drug court: How offenders view ease of program completion, strengths and weaknesses, and the impact on their lives. *National Drug Court Institute Review 2*(1), 61-85.

U.S. Government Accountability Office. (2005, February). *Adult drug courts: Evidence indicates recidivism*

reductions and mixed results for other outcomes. Report to congressional committees. Washington, DC: Author.

White House Office of National Drug Control Policy. (2005). *National Drug Control Strategy*. Washington, DC: Author.

APPENDIX

Table 1. Results of Research Priority Field Review

Research Priority	Mean Response (Standard Deviation)	Percent "Extremely" or "Quite" Important
What are the long-term effects of drug courts and other problem solving courts on important outcomes other than recidivism, such as substance use, psychological health, physical health, employment, or parenting? What components of the drug court model contribute to the most effective outcomes in those areas?	4.48 (0.71)	91%
Do minority sub-groups have differential access to drug-court programs or differential success or failure rates? Are they subjected to different types or amounts of sanctions or rewards for comparable performance? Do they receive different types of treatment services? If so, why and how do we correct this?	3.48 (1.35)	58%
How are outcomes influenced by having a permanently dedicated drug-court judge and docket, as opposed to annually rotating assignments?	4.0 (0.96)	77%
Does it matter whether a judge wants to be in drug court, or can any judge be "brought up to speed"?	3.93 (1.05)	72%
What traits or characteristics of the judge, if any, are associated with better outcomes for various clienteles?	3.63 (1.19)	60%
Can equivalent outcomes be attained using alternative judicial arbiters, such as masters or commissioners? Does this vary by clientele?	2.81 (1.32)	33%

What are the impacts of brief jail sanctions on clients who are noncompliant with their care plans or program requirements? Do these impacts vary by the nature of the clientele? Do they vary by the target behavior?	4.07 (0.96)	83%
Are there sub-groups of drug court clients for which "rational" models of rewards and sanctions are differentially effective or need to be substantially modified, such as mentally ill offenders, juveniles, or psychopaths?	4.04 (0.91)	72%
How are outcomes affected when a drug court imposes a pre-defined "matrix" of sanctions or rewards, as opposed to individualizing its responses and keeping clients "guessing"?	3.98 (1.06)	72%
What is the optimum length of time for required participation in a drug court program? Does this vary by clientele or by the drug of choice? Does drug court "accelerate" recovery because of the additional services and monitoring?	4.19 (0.81)	80%
What are the most effective continuing-care strategies that result in the greatest likelihood of long term success, focusing specifically on practices that (a) utilize the continued influence of the criminal justice system following completion of the drug court program or (b) are embraced by program graduates after the drug court no longer has active jurisdiction over their case?	4.11 (0.88)	80%
Are there sub-types of drug offenders who could benefit from the monitoring components of the drug-court model, even if they are not actively addicted to drugs or alcohol and may not require formal substance abuse treatment?	2.79 (1.06)	27%

What additional or adjunctive services are most related to positive outcomes in drug courts and most likely to serve public-safety aims? In particular, should employment or educational attainments be required prior to graduation from drug courts?	4.04 (0.82)	80%
Are there differential effects when a drug court requires an abstinence-only policy from the outset, as opposed to following a "harm-reduction" approach that approximates abstinence over time?	3.53 (1.08)	56%
What types of clients require frequent judicial contacts, and what types can be effectively and safely managed by community corrections officers, probation or parole officers, case managers, or treatment providers?	3.67 (1.05)	59%
What types of community monitoring technologies and practices are associated with better outcomes in drug courts? Do these outcomes vary by population? How do these technologies and practices affect clients' perceptions of such things as "procedural justice" or "perceived deterrence"?	3.61 (1.03)	61%
How can we develop better methods and instrumentation to measure the degree of collaboration between agencies and systems in drug courts?	3.35 (1.24)	46%
Are outcomes affected by having clinicians and case managers appear during status hearings to give testimony, as opposed to sending written reports or transmitting data elements? Does this affect clients' perceptions of the therapeutic alliance and their willingness to disclose important personal information?	3.49 (1.09)	50%
Are outcomes affected by having clinicians share only limited data elements (e.g., counseling attendance and drug-testing results) with the court and other professionals, as opposed to sharing a wider range of clinical information? Does this affect clients' perceptions of the therapeutic alliance and their willingness to disclose important personal	3.54 (1.06)	58%

information?		
Are decisions more consistent and outcomes more effective when the judge acts as the final arbiter of clients' performance during status hearings, or when the team reaches a general consensus on such matters?	3.90 (1.01)	71%
Does participation in drug court raise the quality of all staff members' performance, such as improving the quality of treatment?	3.17 (1.25)	45%
Are outcomes improved or are services more efficient when the drug court coordinator is an agent of the court system, the treatment system, probation/parole, or some combination of these?	3.27 (1.28)	49%
What methods of client staffing or case-conferences lead to the most effective sharing of information and to the best outcomes? In particular, is it necessary for the prosecution and defense to be present during staffing and during court proceedings?	3.51 (1.04)	52%

DRUG COURT PERFORMANCE MEASUREMENT: SUGGESTIONS FROM THE NATIONAL RESEARCH ADVISORY COMMITTEE

By Cary Heck, Ph.D.
and Meridith H. Thanner, Ph.D.

While drug court research continues to forge a path toward greater understanding of the model, drug court evaluation practices seem to be suffering from a lack of clear direction as to the important elements of drug court programs that should be measured and compared. This paper is an attempt to answer some of the basic evaluation questions for local programs, state drug court management, and federal partners. The content of this article is drawn largely from the work of the National Research Advisory Committee sponsored by NDCI and was motivated by a commonly held belief that many drug court evaluations are in need of guidance (see the 2005 U.S. Government Accountability Office report for examples). The purpose of this paper, therefore, is to promote quality research at all levels for drug court programs by presenting a uniform and manageable data collection and evaluation strategy for local programs. This paper focuses on one element of program evaluation: performance measurement. It provides four essential measures of drug court performance and makes suggestions about how to document and analyze these measures. The presented measures can be used across the spectrum of drug court programs to aid local jurisdictions in answering questions posed by stakeholders and funding agencies, as well as assist in promoting sound management practices at the local court level.

Cary Heck, Ph.D., is an Assistant Professor of Criminal Justice at the University of Wyoming as well as the Director of Research for the National Drug Court Institute. Previously, he served as Statewide Director of Drug Courts for Louisiana through the State Supreme Court. Dr. Heck

has also worked as a small town police officer, a probation officer, a gang task force member, and a counselor for delinquent youth.

Meridith H. Thanner, Ph.D., is a Research Associate with the Bureau of Governmental Research (BGR) at the University of Maryland College Park and an independent consultant with the National Drug Court Institute. Most recently, her work with BGR has focused on managing process and performance evaluation studies of adult DUI courts and juvenile drug courts across the State of Maryland. Other research interests, stemming from her doctoral work in military sociology at the University of Maryland, includes ethnographic assessments of how communities are affected by military base closings and realignments as part of the Federal Base Realignment and Closure Process.

Direct all correspondence to Cary Heck, Ph.D., University of Wyoming, Criminal Justice, Department 3197, 1000 East University Avenue, Laramie, Wyoming 82071. (307) 766-2614; check@ndci.org.

ARTICLE SUMMARIES

WHAT IS PERFORMANCE MEASUREMENT?

[5] Performance measurement refers to the establishment of research-based indicators to measure program activity. Evaluation in this form allows for program feedback as well as cross-site comparisons.

MEASURING DRUG COURT PERFORMANCE

[6] There are four measures of drug court performance recommended: retention, sobriety, recidivism, and units of service. All of these measures can be examined at either the client level or the program level.

CONCLUSION

[7] While not a replacement for a focused process evaluation, performance measurement can help establish a basis for funding and implementation decisions as well as bringing greater state and national representation to local programs.

INTRODUCTION

One of the most effective ways to tout the benefits of adult drug court programs, as well as to silence critics, is to show positive client outcomes based on rigorous data collection methods and sound analysis. Drug court research and evaluation, however has had trouble keeping up with the rapid implementation of drug court programs, despite many state and federal agencies including evaluation as part of their funding requirements. The task of evaluation has proven to be a challenge for many local jurisdictions struggling to sustain their programs with a sparse number of team members and limited financial support. This is particularly the case since most drug court teams do not have an evaluation component built into their daily operations model, either with the assignment of data collection and management tasks to a member of the team for evaluation purposes, or the hiring of an outside evaluator to oversee such tasks on an on-going basis. Of the drug court programs that do collect and analyze data related to client and program performance (either in-house or with the assistance of an outside evaluation team), there is a lack of consistent and uniform method across sites. More importantly, many of the evaluations conducted are not methodologically sound and thus are not able to abate continued skepticism regarding the effectiveness of drug courts.

Worse yet, few programs seem to understand the importance of adequate data collection at the inception of the program, making them poor candidates for evaluation services. These problems became particularly obvious during the Evaluation Plan Review Project undertaken by the National Drug Court Institute's (NDCI) research team in 2004, in which more than 100 evaluation plans submitted as a special requirement of federal drug court implementation grants were reviewed. Through that project, it became very clear to the review board that local drug courts need guidance on how best to evaluate and report their activities.

The purpose if this article, therefore, is to assist program managers, state leaders, and evaluators in developing clear and meaningful evaluation plans that truly reflect program activities through uniform performance measures. When done correctly, these measures can be compiled at the local level and aggregated at the state and federal levels. Additionally, the compilation of these measures will provide a scientifically sound means for comparing drug court program performance both between programs and to other interventions.

The National Research Advisory Committee (NRAC) was formed to develop guidelines for program evaluation and performance measurement. This committee is comprised of many of the leading research scholars in drug court field. Over the course of three meetings, the committee compiled its recommendations, some of which are reflected and summarized in this article. During this process it became apparent that measurement of drug court programs needed to be clear, succinct, and manageable. Thus, a *performance measurement* model was selected.

WHAT IS PERFORMANCE MEASUREMENT?

[5] Performance measurement is an excellent option for drug court research and can assist in developing correlations between program activities and outcomes. Performance measurement refers to the establishment of research-based indicators to measure program activity (Epstein, Coates, Wray, & Swain, 2005). There are several performance measures for drug courts that might be used to effectively document the effects of drug courts on clients. Four of these measures will be discussed below. However, some ground rules must be established for the use of these evaluation tools.

A great deal of confusion surrounds the constructs of outcome/impact evaluations and performance measurement.

Outcome and impact evaluation both imply determining a *causal* relationship between a program or policy and some greater social gain or loss (Fitzgerald & Cox, 1994). True outcome evaluation requires the use of an experimental design and randomized selection of participants. However, the political, judicial, and social arenas in which drug courts operate make it unlikely that many studies will be able to employ randomized subject selection. More commonly, local evaluation projects can establish *correlations* between drug court program practices and the intended consequences of the intervention through methodologically sound evaluations that compare the individual impacts of drug court participation on clients as compared to those without this intervention (Johnson & Wallace, 2004).

Since the delicate relationship between *causation* and *correlation* can easily be confused, it is incumbent upon drug court researchers to be very clear which they mean. To declare *causation*, randomized control groups must be developed to measure the effects of drug courts on clients in comparison to traditional business-as-usual methods (Fitzgerald & Cox, 1994). Pure causal inference cannot be drawn without random assignment of subjects (King, Keohane, & Verba, 1994). An excellent example of causal research in drug courts is the recent work by Douglas Marlowe and his colleagues in a study of Delaware drug courts (2004). Using random assignment, drug court participants were assigned to one of two groups. The first of these groups was required to attend bi-weekly judicial status hearings regardless of their program performance. The second group only had to attend judicial status hearing on an "as needed" basis; for participants in this group, status hearing attendance depended upon their behavior in the program. Given the random assignment of this model, researchers were able to make *causal* claims about the impacts of these hearings (Marlowe, Festinger, & Lee, 2004).

Correlation, however, is mainly concerned with the strength of the relationship between two variables. Variables are said to correlate if a change in one variable influences a second variable. For example, an evaluation of the Chester County Drug Court Program in Pennsylvania compared program participants to a matched sample of offenders that did not participate in drug court on the question of in-program recidivism. Using this model, the author was able to claim that those participating in drug courts had lower recidivism than those in the comparison group (Brewster, 2001). Thus, a correlation was drawn between drug court participation and offender recidivism. The relationship between the drug court program and client behaviors, as investigated by most drug court evaluations, can be said to be correlated, not causal.

Experimental design is the time-honored and proven way to discover the effect of a treatment on a population. Its fundamental tenet is the use of a control group—the randomly assigned group that does *not* receive the treatment, providing a non-treatment group that the experimental group can be measured against. The world of criminal justice practice, however, is not a laboratory. As a result, the ability to use control groups, and thus an experimental design, is severely compromised. Quasi-experimental design, then, is the next best option. As its name suggests, quasi-experimental design is *almost* experimental, and therefore its findings *almost* as credible (Campbell & Stanley, 1963). Instead of control groups, this design uses comparison groups, which can provide information that is both useful and important. These groups are made up of individuals that mirror those being studied in important ways. In "matched groups," as they are sometimes called, the group is matched to the experimental groups on important variables, sometimes individually. For example, it would make sense to compare drug court clients to criminal offenders with substance abuse problems, but depending on the target population, it may not

make sense to include violent offenders in the comparison group.

While comparison groups do not eliminate the problem of selection bias in research, they make possible reductions in the likelihood of selection bias by increasing the points of comparison. If drug court researchers are interested in comparing drug court clients to those who do not receive drug court intervention, it is important to determine the important personal variables that might lead to program success or failure. In a more detailed example, drug court participants can be compared to non-drug court participants through the creation of comparison groups using official data sources (such as local arrest databases or judicial records). Depending on the data source available, two different types of comparison groups can be created—a historical home comparison group and/or a contiguous community comparison group. The historical home comparison group is comprised of individuals from the same jurisdiction as the treatment group who would have been eligible for the drug court program had the program been implemented at the time of their involvement with the judicial system. Thus, this group represents the same geographical area, but a different time period. To create this group, researchers can use records from approximately 12 to 24 months prior to the implementation date of the drug court to identify individuals who meet the eligibility criteria of the program and who, as a group, are statistically comparable to the treatment group on key variables.

On the other hand, the contiguous community comparison group is comprised of individuals from a contiguous community who would be eligible for the drug court program, if there was a drug court program in their jurisdiction. Thus, this group represents the same time period, but a different geographical area. To create this group, researchers can use records from the same time period as the treatment group, but from the contiguous non-drug

court program community, to identify individuals who meet the eligibility criteria of the program and who, as a group, are statistically comparable to the treatment group on key variables. Analysis, therefore, includes comparisons of outcomes of the treatment group to the comparison group (either the historical home group or the contiguous community group) to identify the impact of participation in the drug court program on the likelihood of, for example, recidivism or sobriety in comparison to those not participating in the program.

One evaluative mistake that is often made when creating a comparison group is outcome comparisons (i.e., recidivism) between program graduates and terminations or failures. Despite the allure of what appears to be a convenient comparison group, use of program dropouts and absconders as a comparison group *is not valid science*. Any outcome comparison must be done between the entire drug court participant group, *inclusive of failures*, and another entirely separate group. To compare failures to successes in this manner is very much akin to comparing high school students with straight A's to those students with F's. Most schools have "A" students that can make them look good—the question is how good the overall quality of education is for *all* students.

However, the comparison of dropouts to graduates for other reasons can bear interesting results and should be encouraged. Comparing these two groups on matters such as program satisfaction, cultural competency, or treatment participation may yield findings that could assist a jurisdiction in ultimately achieving a better graduation rate. Conversely, examining the two groups for glaring differences could also provide some insight into other areas for study and correction. If dropouts seem to be predominantly female, minorities, or of a certain age, it is possible that additional study could reveal weaknesses in the court's treatment of these populations.

What Makes Performance Measurement Valuable?

Standardized performance measurement can provide program managers and those with funding authority a means by which to accomplish two things. The first is program feedback. Program feedback allows managers the ability to identify areas of strength and weakness and seek solutions to problem areas (Bachman & Schutt, 2003). For example, several programs around the country have identified problems with retention of certain sub-populations, particularly those who are young or have co-occurring disorders. Using a standardized tool to measure retention makes it possible to perform analysis among groups within the programs and then implement programmatic changes designed to address the problem.

Secondly, standardized performance measures allow for the compiling of data across programs and locations. And, while it is always important to consider context when viewing these numbers, oversight is enhanced and the numbers are readily useable for reporting program activity to funding agencies. For drug courts, this process will most often occur at the state level. These performance measures also promote accurate and consistent reporting to legislative bodies that are generally charged with allocating program funding.

MEASURING DRUG COURT PERFORMANCE

[6] In the interest of uniformity and with a realistic understanding of the research capacity of local programs, NRAC chose to focus on three primary and one secondary measure of program performance. Retention, sobriety, and recidivism cover a great number of important effects of drug court, and as such, should serve as the primary performance measures. Units of service also should be considered a secondary measure of program performance. These measures are described in this section with suggested definitions and

measurement strategies. It should be noted that these modes of measurement are not the only means for documenting drug court activity, and the definitions are not yet universally accepted. Regardless, it is incumbent on drug courts to document program performance in a manner that can be compiled and compared.

Retention

Retention has often proved difficult for drug court professionals to calculate. It should be calculated as a ratio or percentage: the retention rate is the number of people that complete or remain in the program divided by the number that enter the program during a particular time period. Thus, all of those who depart the drug court for any reason, including, but not limited to, those who abscond, voluntarily withdraw, and are expelled, should be included as part of the denominator. However, it is impossible to calculate retention without considering drug court clients as a *cohort*. Overall program retention should be the ratio of those who complete the program or are still enrolled in the program divided by those who enter the program during the time frame under consideration, generally six months to one year. Since some participants who are still enrolled when a court decides to assess retention may ultimately drop out, the retention rate may need to be recalculated once the entire cohort has departed the drug court, either successfully or unsuccessfully.

A *cohort* is a group of individuals who enter the program during a particular time period. The court can define the time period, depending on the number of clients served in the program. Generally, a 6-month or 1-year time period is considered appropriate for developing a drug court cohort. For example, a court operates with an average of 100 total clients. The program requires clients to complete 12 months of continuous participation in treatment and court activities. Fifty clients entered the program during the first 6 months of 2005; this is defined as the *retention cohort*. At the end of

the first 6 months of 2006, a retention rate could be calculated using 50 as the denominator. In this case, 5 clients opted out of the program and 5 more were dismissed from the program, leaving 40 clients from that 6-month period that eventually graduated (even if it took longer than 12 months to graduate) or are still in the program. The retention rate would then be 40/50, or 80 percent.

Sobriety

Documenting the continuous sobriety of drug court clients is one of the highlights of any drug court evaluation. Sobriety is most reliably measured using clean drug screens. Best practices for drug court suggest frequent and random screens. Self-reported drug use during the program without a formal drug screen result is not considered a reliable measure. All drug screens and the results thereof, both positive and negative, should be documented, as well as those that are missed, excused, tampered, stalled, or inconclusive. Missing and tainted drug screens should be counted as dirty and should break the chain of continuous abstinence. In this way, it will be possible to develop and record benchmarks for clients. Overall program performance can be documented using average length of sobriety during a specific timeframe. Drug courts should be able to document both the *average* length of continuous sobriety and the *average* number of failed tests that a client has during the program or during a particular time period. Theoretically, a trend should exist among drug court clients demonstrating reduction in the number of dirty drug screens over the course of the program. Trends can be documented by compiling information from clients over time. Both the trend and the averages will prove useful measures of drug court performance.

Recidivism

Recidivism has traditionally been a contentious subject. The term simply means a return to criminal activity

by someone who has already been adjudicated guilty or delinquent, but the difficulty for some researchers comes in the attempt to measure the concept. For the purposes of drug court research, it is suggested that drug court evaluations use arrest as the primary measure. This choice reflects several factors, including ease of documentation, as well as the accelerated turnaround time for processing documentation not found with other methods commonly used, such as conviction. Maintaining records of both measures could prove highly useful for research purposes, but the ramifications of conviction render it less useful than arrest for evaluation purposes. Often, clients who are charged with additional crimes plead out or are given other diversionary programs that prolong the process. In considering in-program recidivism, researchers should remember that it is much more likely that clients will be arrested and charged with a crime during the program than will actually be convicted. Therefore, arrest is a better measure for evaluation purposes.

To the extent possible, it also is valuable to collect conviction data. Simply put, conviction data are related to the extent to which those who were arrested for subsequent offenses were charged and convicted of these crimes. There is much debate about whether arrest data or conviction data are accurate measures of criminality. Both of these can proxy measures of recidivism; although neither is perfect, there is definitely a need to report such indicators of program performance to paint as full a picture as possible, and as such, conviction can serve to augment arrest data.

Recidivism also is the one performance measure that could plausibly be considered after program completion. It is recommended that, to the extent possible, programs develop methods to track clients after program participation to examine this, using information from the local justice process as well as state and National Crime Information Center (NCIC) databases. The use of a comparison group enhances this type of research, but the data can be useful on its own.

Using post program recidivism data, researchers can make some claims about the impact of the program on client behaviors. This model should allow drug courts to build on the sample data collected by the National Institute of Justice and the Urban Institute (Roman, Townsend, & Bhati, 2003) to describe drug court recidivism in a more complete way.

Units of Service

Units of service can be loosely defined as a measure of those drug court activities that address the needs of drug court clients, including, but not limited to, substance abuse treatment. These measures of drug court performance are easy to neglect when considered alongside more obvious issues like recidivism and retention rates. It is, however, vital that all activities of court programs be documented for two reasons. First, drug court program managers need to determine which services are affecting the clients in a positive way. In doing so, managers and judges can evaluate the efficacy of the various interventions used to benefit clients. Second, and perhaps of greater importance, is the need to both display and fully understand the brokerage of services and the collaborative nature of drug courts that are their major innovation from traditional judicial practices. Many programs provide medical, mental health, vocational, and educational programs for clients beyond the standard drug treatment. For some clients, these services may be at least as important as the treatment itself.

The use of a "unit of service" modality for measuring drug court activity is a simple means for documenting these secondary court activities. Service units should be based on the actual attendance of a drug court client in one of the recommended or mandated activities. Unit of service measurement must go beyond referral, although it is valuable to track this as well. If a client were remanded to a job-training program and attended three 1-hour classes per week, each class could be considered a service unit. Likewise, a

visit by a client to a psychiatrist to treat a co-occurring disorder would be counted as a service unit. Outside assessments and consultations also should be documented. Often, billing sheets can assist in tracking services. Inpatient treatment is most easily considered using "days" as the measure of a service unit.

Client-Level and Program-Level Variables

These variables can be considered as both "client-level" and "program-level" variables. Client-level variables refer to those variables related to a particular client. Thus, using the performance measures listed above, programs can look at individual client performance in the areas of retention, sobriety, recidivism, and services received. However, it is also important to look at these variables from a program level. That is, it is useful to look at program performance by compiling the numbers related to the client level variables on the four important dimensions. An example was given above on how to translate an individual-level variable (i.e., program retention) into a program-level variable. Using a similar mode of calculation, programs can determine the average length of sobriety measured in days, the recidivism rate, and the average units of service provided for clients.

Some Caveats

There are at least two important caveats that must be mentioned in relation to performance measurement and drug court. The first is that drug court programs, while similar to one another in many ways, have differences that must be considered when comparing performance. These differences often include issues related to the population being served and the availability of resources. Drug courts that serve younger populations should, according to the research, have poorer outcomes. Likewise, those that are limited in their treatment capabilities (e.g., no inpatient treatment services available) might also exhibit lower performance.

Secondly, programs in varying stages of implementation will inevitably have different levels of performance. As programs become more established, they tend to find a niche that supports improved programmatic outcomes and performance. Performance measurement can be used to provide support for program improvements but researchers should not forget the program's stage of development when considering these measures in relation to other programs.

CONCLUSION

[7] Performance measurement provides a strong tool for program managers to document program outcomes and define areas that might need improvement. The four performance measures presented and described in this paper can serve as a solid research foundation for local programs and an excellent means for compiling data at the state and federal levels of government. Each of the measures was carefully and thoughtfully considered by the members of NRAC and determined to be adequate for the broad documentation of drug court activities. However, they should not be considered a replacement for in-depth process evaluation of drug court programs. Rather, they should serve as part of a regular and on-going review of drug court programs.

Stakeholders, including those with decision-making authority regarding funding, should find these measures to be adequate for establishing a basis for funding and implementation decisions. These drug court performance measures are not meant to replace experimental designs for research but to serve as a meaningful and practical means to evaluate of the drug court performance. It is exciting to consider how drug courts can be represented on a state and national level if these measures are accurately gathered and compiled.

REFERENCES

Bachman, R., & R.K. Schutt. (2003). *The practice of research in criminology and criminal justice* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Brewster, M.P. (2001). An evaluation of the Chester (PA) drug court program. *Journal of Drug Issues*, 31, 177-206.

Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.

Epstein, P.D., Coates, P.M., Wray, L.D., & Swain, D. (2005). *Results that matter*: *Improving communities by engaging citizens, measuring performance and getting things done*. San Francisco: Jossey-Bass Publishing.

Fitzgerald, J.D. & Cox, S.M. (1994). *Research methods in criminal justice: An introduction*. Chicago: Nelson-Hall Publishing.

Johnson, C.M., & Wallace, S. (2004). Critical elements to consider for methodologically sound impact evaluations of drug court programs. *Drug Court Review IV*(2), 35-48.

King, G., Keohane, R.O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.

Marlowe, D.B., Festinger, D.S., & Lee, P.A. (2004). The judge is a key component of drug court. *Drug Court Review IV*(2), 1-34.

Roman, J., Townsend, W., & Bhati, A. (2003, July). *National estimates of drug court recidivism rates*. Washington, DC: National Institute of Justice, U.S. Department of Justice.

U.S. Government Accountability Office. (2005, February). *Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes*. Report to congressional committees. Washington, DC: Author.

EVALUATING DRUG COURTS: A MODEL FOR PROCESS EVALUATION

By Cary Heck, Ph.D., and Meridith H. Thanner, Ph.D.

Process evaluations are important tools for program management and oversight. Done well, drug court process evaluations should provide program managers with insight into their program's operations as they relate to the fundamental mission of improving the long-term prospects for their clients. Additionally, process evaluations of drug courts should promote consistent data collection and analysis of drug court activities. This article discusses and presents the elements and realities of process evaluation with the aim of assisting local programs in working with independent evaluators to develop and sustain ongoing process evaluation mechanisms. It is largely the product of the NDCI's National Research Advisory Committee and thus is the synthesis of suggestions from a broad base of well-known substance abuse and drug court researchers.

Cary Heck, Ph.D., is an Assistant Professor of Criminal Justice at the University of Wyoming as well as the Director of Research for the National Drug Court Institute. Previously, he served as Statewide Director of Drug Courts for Louisiana through the State Supreme Court. Dr. Heck has also worked as a small town police officer, a probation officer, a gang task force member, and a counselor for delinquent youth.

Meridith H. Thanner, Ph.D., is a Research Associate with the Bureau of Governmental Research (BGR) at the University of Maryland College Park and an independent consultant with the National Drug Court Institute. Most recently, her work with BGR has focused on managing process and performance evaluation studies of adult DUI courts and juvenile drug courts across the State of Maryland. Other research interests, stemming from her doctoral work in

military sociology at the University of Maryland, includes ethnographic assessments of how communities are affected by military base closings and realignments as part of the Federal Base Realignment and Closure Process.

Direct all correspondence to Cary Heck, Ph.D., University of Wyoming, Criminal Justice, Department 3197, 1000 East University Avenue, Laramie, Wyoming 82071. (307) 766-2614; check@ndci.org.

ARTICLE SUMMARIES

WHAT ARE PROCESS EVALUATIONS?

[8] Process evaluations are tools that drug courts should use to measure their efficiency, efficacy, and achievement of program goals.

WHO SHOULD CONDUCT THESE EVALUATIONS?

[9] Trained evaluators have a skill set specific to the task and should be willing to listen to staff concerns.

WHAT ARE THE CRITICAL ELEMENTS OF PROCESS EVALUATION?

[10] Critical elements of analysis include program goals, target population, drug treatment, court processes, units of service, team cooperation, and community support.

WHAT DATA ARE NEEDED TO COMPLETE THESE EVALUATIONS?

[11] Data collected must be valid and reliable over time. Confidentiality guidelines must be followed in the collection of sensitive data.

WHAT DOES IT MEAN TO BE "METHODOLOGICALLY RIGOROUS"?

[12] Even though experimental design may not be feasible, evaluation research must still follow accepted guidelines in its methodology.

WHAT ABOUT EXPERIMENTAL DESIGN AND COMPARISON GROUPS?

[13] Quasi-experimental design is the most practical method of evaluating program outcomes and impacts. Comparison groups should be matched to drug court groups on the basis of research-established factors.

INTRODUCTION

Drug courts are a national phenomenon. Few, if any, criminal justice interventions have spread throughout the country with as much speed and support as drug court programs. Drug court programs started with a single locally driven project in 1989 and have grown exponentially in number since. In 2004, the total number of operational drug court programs in the United States reached 1,621 and the total number of problem solving courts (e.g., mental health courts, domestic violence courts, family treatment courts) in general was 2,557 (Huddleston, Freeman-Wilson, Marlowe, & Roussell, 2005). The reason for this rapid growth is three-fold. First, drug court programs are based on the intuitive model of program design and implementation; that is, this person-centered model requires relatively little investment in order to realize long-term public support savings. Thus, the broad societal benefits of implementing a drug court program far out weigh the incremental costs involved in its formation and maintenance. Second, in light of the noticeable successes of drug court programs at the local level, a tremendous word of mouth public relations campaign has been undertaken by judges, drug court professionals, and clients, which has led to an impressive number of anecdotal cases in support of the model. Third, there is a growing body of empirical research that supports drug courts as effective programs for dealing with substance abusing offenders. Drug courts are credited with reducing recidivism, retaining clients in treatment, and improving outcomes and quality of life circumstances for substance abusing offenders. And, as evidenced in an increasing body of scientifically rigorous research, it is clear that drug courts are effective alternatives to traditional "business as usual" methods (Marlowe, DeMatteo, & Festinger, 2003).

The Current State of Local Drug Court Evaluation

While the growth of drug court programs and the increasing body of literature suggesting the effectiveness of drug courts are undeniable, the quality and utility of many local program evaluations and data collection strategies remain questionable. Drug courts have historically faced considerable criticism in the area of evaluation and documentation. Through the Bureau of Justice Assistance (BJA), the federal government allocates millions of dollars to fund local drug court programs, but despite repeated efforts to count and document the activities of these programs, there is little uniform data on actual drug court activities nationwide. Congress has asked the U.S. Government Accountability Office (GAO) to review drug court evaluation and outcome research to determine the effectiveness of drug courts no less than four times. The most recent GAO Congressional Report on adult drug courts only found 27 of 117 evaluations of local drug court programs to be of sufficient methodologically quality to use for analysis (GAO, 2005). The findings from these evaluations indicate uniformly that drug courts produce positive results, but the lack of a broader selection of methodologically sound evaluations has led to continued skepticism.

In 2004, the National Drug Court Institute (NDCI) entered into an agreement with BJA and the National Institute of Justice (NIJ) to review evaluation plans for all federal drug court grantees. These evaluation plans were a special condition for grant recipients. NDCI reviewed over 100 evaluation plans and found considerable variation in the quality of the proposed research. Many of the evaluation plans required considerable technical assistance and a few were considered to be completely without merit and required rewriting. Programs often had difficulty identifying an appropriate and rigorous evaluation method, little theoretical understanding to guide their proposed evaluation activities, and poor data collection systems to support the conduct of a

methodologically sound process and performance evaluation. Furthermore, anecdotal evidence suggests that some methodologically sound evaluation plans are never carried out.

This relative lack of quality evaluation research has created difficulties for the national drug court movement and local programs alike. At various times, scholars and politicians have questioned both the effectiveness and efficacy of drug court programs and some continue to argue that drug courts are not worth the money being spent. While local drug court programs continue to build support and thrive, the federal resource allocation to drug courts is open to challenge and the movement has limited solid evaluation research to refute its critics. Further, no less than 35 states have appropriated funding for drug court programs. While some of these appropriations are pass-through funds from federal programs, others are direct legislative appropriations from general funds, and many are a combination of the two. A recent calculation of state appropriations for drug courts totals close to $150 million annually (Huddleston et al., 2005). To some degree, all of these appropriations are dependent upon the ability of local programs to document their work and report their outcomes.

A National Strategy

The purpose of this paper, therefore, is to promote quality research at all levels for drug courts by providing a uniform and manageable process evaluation strategy for local programs (see the other articles in this issue for information on recidivism and performance measurement). These methods can be used across the spectrum of drug court programs to allow local jurisdictions to answer questions from stakeholders and funding agencies, as well as promote sound management practices at the local court level. It is clear that research practices can be improved by providing a uniform baseline for evaluation and measurement. This paper

concentrates primarily on *adult* drug courts. While many of the ideas can easily translate to other problem solving courts, the scope of this article is limited to promote research accuracy. Adult drug courts are an appropriate focus mainly due to their prevalence; NDCI places the number of adult drug courts in 2004 at 811 nationally (Huddleston et al., 2005).

It should also be noted that the methods suggested in this paper are not the only ways to gather useful information regarding drug court processes and performance—local court programs need to direct their own research to benefit their own programs. Furthermore, it should be understood that the quality of research depends heavily on access to data and the availability of resources. With this in mind, some additional suggestions for improving research design beyond the baseline requirements are found in the following pages.

This paper focuses on six important questions related to drug court evaluation. The first two questions deal with the general elements of process evaluations and the issue of who should conduct these evaluations. The third and fourth questions relate to the actual performance of local drug court program evaluation, including a discussion of the critical elements of conducting this type of evaluation as well as what data are needed for the purpose of answering these questions. The fifth and sixth questions relate to the issues of methodological rigor and evaluation design.

WHAT ARE PROCESS EVALUATIONS?

[8] Generally speaking, evaluation research refers to a purpose rather than a specific methodology (Maxfield & Babbie, 2005). Simply put, process evaluations should be tools for managers and stakeholders as they seek to maintain successful programs, enhance services, and promote research-based best practices within programs (Rossi & Freeman, 1989). Evaluation research is a means by which programs

can be opened up to determine the extent that they are achieving program goals and managing their activities in an effective and efficient manner. Likewise, the intended audience must motivate process evaluation. In many cases, certain audience members (i.e., state administrators) will ask specific questions that evaluators can address in their reports. Two primary questions drive process evaluation: policy implementation and the achievement of program goals (Maxfield & Babbie, 2005).

Further, evaluation research is commonly defined using three important constructs. The first is the use of a systematic approach in synthesizing evaluation plans (Rossi & Freeman, 1989). That is, the plan must be designed and implemented in a strategic, careful, and consistent manner. This is particularly important since the evaluation often relies on retrospective data, which may or may not have been collected and managed in an easily accessible format. This systematic approach applies not only to the collection of information, but also to the second major construct, the critical analysis of information (Rossi & Freeman, 1989). It is clear that simply collecting information is insufficient to constitute a valid evaluation; it is critical that the information be carefully *analyzed* by individuals or teams that understand the underlying principles guiding program practice. Finally, evaluation research must provide *useful* feedback. Evaluative feedback that is difficult to understand or meaningless to the consumers serves little purpose for the program, and thus is not practically useful, though perhaps highly advanced and descriptive (Bachman & Schutt, 2003).

Defining Terms

With this in mind, it is important to discuss some basic evaluation research terminology. Evaluators generally consider four *terms of art* when developing evaluation methodologies (Bachman & Schutt, 2003). The first of these terms is *inputs*. Inputs can be considered to be any of the raw

materials that enter the program. For example, drug court inputs tend to be clients, program staff, and additional resources. The second term, *program process*, refers to the treatment and/or services provided for clients in the program, as well as the policies that guide the delivery of those services. Drug courts rely on a variety of process mechanisms to create positive effects including sanctions and incentives, substance abuse treatment, and ancillary services. Third, *outputs* are short-term products produced by the program process. For drug courts, this term could be used to refer to the number of hours of substance abuse treatment received by a client or the number of urine screens that a client provides. Finally, *outcomes* are the impacts that the program has on its participants (Bachman & Schutt, 2003). Often the terms *outcome evaluation* and *impact evaluation* are used interchangeably. However, there is at least one subtle difference between the two: Impact evaluations tend to focus on large scale measures of quality of life beyond the particular client, while outcome evaluations tend to focus on the effects of the program or policy on a particular participant.

Measurement

Process evaluators must consider each of these evaluation terms (or constructs) and find appropriate means for measuring them within the context of the program. Measures that are too broad or too narrow often tend to over or underestimate program effects. It is critical, therefore, that evaluators carefully consider the variables used to measure client background and risk, program activity, outputs, and outcomes. The best means for identifying the important variables is through a process that first defines the questions and then links the questions to variables that can be sufficiently applied based on existing research or a researchable hypothesis. Recent reviews of drug court evaluation plans by the NDCI research team revealed some significant uncertainty about how best to measure the

important evaluation questions facing drug courts. Often the questions were appropriate but the means identified for answering them were unexplained, unclear, or inappropriate. For example, many evaluation plans did not consider previous treatment failures as an important client characteristic for evaluating their target population. Many simply considered legal measures (i.e., criminal history) as the only defining social variable apart from simple demographics. Existing research, however, clearly shows that certain drug court models and activities perform better with particular types of substance abusing offenders than do others (Marlowe, Festinger, & Lee, 2004).

Operationalization is the term used by researchers to define the process of making a construct measurable or turning concepts into variables. It is the act of taking a term like "recidivism" and making it measurable and comparable. It requires an understanding of the definition of the term (i.e., offender committing an additional criminal act after being arrested, charged, or convicted for a criminal act) and creating a meaningful way to measure it (e.g., arrest on a new charge). Often the operationalization of a construct is not exact, as demonstrated by the example above. However, it is important that the method used for measuring the construct be theoretically defensible given the context. The recently completed monograph by NDCI's National Research Advisory Committee, *Local Drug Court Research: Navigating Performance Measures and Process Evaluations*, suggests using arrest data for analysis of recidivism (Heck, in press). Clearly, arrest data have weaknesses as measures of actual criminality but, given the theoretical defensibility of the choice, constraints of data collection, and the length of time required to get through general court proceedings, it was decided by the committee that this measure would be the best for the purposes of performance measurement.

Local Drug Court Process Evaluations

Drug court process evaluations are tools to be used by programs for improvement and should provide interested parties with a glimpse into the workings of a drug court program—specifically, it should elucidate how the operations of the court produce its effect. These evaluations are focused upon the *how* and *why* of drug court activity. Minimally, a process evaluation should include fundamental descriptive statistics (e.g., simple summaries of certain samples or measures such as the number of men and women in the program, the number of court appearances, the modality of treatment offered) and use these to answer questions concerning the level to which programs are meeting their operational and administrative goals. One common process question focuses on the extent to which the local program is reaching the population it was chartered to serve. By definition, drug courts target particular types of offenders. Those eligible often include offenders with no prior violent history and substantial addiction problems. After reviewing the program's target population goals and comparing this to the type of offender the program is actually accepting, an evaluator may, for example, suggest the refinement of the program's target population, or a refinement in the decision-making process used to accept certain clients in order to make better use of limited resources. By focusing on evaluating the target population, a court is able to better understand its own screening process, as well as evaluate the suitability of its ideal client group, given the resources available.

A comprehensive evaluation of a drug court program should take into consideration the structure and process of the program in addition to examining program impacts (e.g., participant outcomes). Ideally, evaluations should examine how program structure and process contribute to found impacts. In this way, evaluators can help programs answer not only *whether* the drug court works, but also *how* the drug court works. Moreover, evaluators can help program

administrators understand how the internal functions of the court affect client behaviors during their time in the program and beyond (Goldkamp, White, & Robinson, 2001; Longshore et al., 2001). Due to several factors, including a lack of resources, drug court evaluations often only report on client outcomes and do not include a description, discussion, or analysis of important and contributing program process elements. Further, these reports often focus upon outcomes and measures that are unrelated to the program's goals. Given that drug courts operate as a function of local interests, needs, and resources, the lack of attention to (or at least presentation of) court process elements in these very localized programs has also hindered the drug court movement as a whole, as well as the development of a broad and comprehensive process evaluation model that could help guide programs and their evaluation efforts. Though drug court professionals and practitioners have recognized this as an important function to the future sustainability of the drug court movement, the lack of education, training, and technical assistance on the mechanics of conducting an appropriate and methodologically sound process evaluation has often stalled this undertaking (Office of Justice Programs, 1998). This paper provides a resource for addressing this issue.

WHO SHOULD CONDUCT THESE EVALUATIONS?

[9] Process evaluations must be conducted by objective outsiders with knowledge specific to the area in question. Independent evaluators are less susceptible to political and personal pressures while conducting evaluations. Though many programs use self-evaluation models and have program employees conduct their own evaluations, the multiple purposes of program evaluation are not well-served by these methods. As such, we recommend two primary considerations when identifying a prospective evaluator. First, the evaluator must understand evaluation. There is a widespread assumption that anyone with a higher education degree has a fundamental understanding of program

evaluation. This is simply not true. Some disciplines and educational programs focus on evaluation research while others do not. Further, an understanding of scientific methods, though helpful, does not in itself imply the mastery of the skills needed to actually conduct an evaluation. Program evaluation is a specific skill. While many of the evaluation plans reviewed by NDCI had quality researchers guiding them, the proposed design was often inappropriate for drug court evaluation.

One way to ensure that potential evaluators have the requisite skill-set is to look carefully at their research background. Drug court program managers should ask potential evaluators to share previous research project reports, and inquire about the methods that they would employ for this type of research. While experimental designs provide for excellent research, they are generally not appropriate or necessary for local drug court program evaluations. Program evaluators must be skilled at researching the program as it operates in the real world. This means that evaluators must be cognizant of context and program limitations given the social and political environments in which they exist.

Second, program evaluators should have some substantive knowledge of the fields in which the program operates. For drug courts, this means that evaluators with experience in substance abuse treatment, corrections, and court processes are preferable to those without, or with a background in just one of these areas. And, while it is theoretically possible for evaluators to become familiar with these substantive areas while working on the project, bringing the evaluator "up to speed" in these areas would involve a great deal of time and energy that might be better expended elsewhere. Additionally, the academic knowledge of subjects such as behaviorism and substance abuse treatment (core elements of drug courts) is not easily gained in the short term.

While self-evaluations can be useful for program management purposes, these types of evaluations rarely yield the comprehensive and rigorous insights that can be articulated by a trained evaluator. Many such "self-evaluation" models are simply means by which employees collect their thoughts concerning the program in a uniform way. It would, of course, be incorrect to suggest that this modality could result in no possible positive results. However, as mentioned above, many consider employees who have a stake in their own programs to be less objective when it comes to the concerns of their programs. Simply put, outside stakeholders will often consider these evaluations to be less credible than those conducted by an outside, objective researcher. Importantly, it is often the case that program employees "cannot see the forest for the trees," and thus lack the ability to view the program in context. Moreover, program evaluators with pertinent experience can often provide ideas and strategies for improving the effectiveness of programs that often go unconsidered by program staff.

Finally, it is strongly suggested that program managers consider the evaluator's willingness to listen to staff concerns during the selection process. Regardless of the amount of programmatic experience the researcher brings to the discussion, managers must remember that evaluations can be guided to answer specific questions that might plague programs. Evaluators should provide systematic analysis of all of the aspects of drug court program operations. Beyond the basic elements inherent to the drug court process, program managers must direct evaluators to consider questions specific to their jurisdiction. For example, if the process of acquiring new, appropriate clients moves at a pace slower than is optimal, the program manager might share his or her concerns with the evaluator and request particular attention be paid to the topic. In general, process evaluations should provide managers with useful feedback regarding the form and function of their programs, with the intent that this

information guide appropriate program improvements, as well as help to document program quality.

To do this, it is important that process evaluators have a solid understanding of the academic research related to drug courts, addiction, and treatment, and be willing to listen. Process evaluations should be conducted with substantial consideration given to the environment in which a drug court program operates, including the actual day-to-day operations of the court, as well as the theoretical constructs associated with the growing body of literature surrounding substance abuse treatment. Sound process evaluations should provide information that is not only based in the research literature, but that is also practical and locally relevant.

WHAT ARE THE CRITICAL ELEMENTS OF PROCESS EVALUATION?

[10] The drug court model has been well defined since the Office of Justice Programs (OJP) and the National Association of Drug Court Professionals (NADCP) released the monograph entitled "Defining Drug Courts: The Key Components" (NADCP, 1997). In fact, many states have copied the key components in their enacting legislation. This model has been utilized with success for over 17 years and while there remain some questions about what parts of the model are most effective (see Goldkamp et al., 2001), there is considerable evidence to suggest that the totality of the approach is effective for retaining clients in treatment and promoting positive outcomes. Thus, one of the important evaluation questions must focus on the integrity of the model as applied by the particular program. These key components are not difficult to operationalize and as such should be included in program evaluations. While it may be the case that variations from this model have developed over time that provide improved services for drug court clients, these variations themselves can be valuable lessons that should be discussed in the analysis of the active program.

At a minimum, there are some basic elements that should be considered in any systematic process evaluation of drug courts. By definition, the following elements should be common to *all* drug court programs:

1. **Program Goals** – Drug court evaluators should examine the extent to which programs are meeting their stated or written goals. Suggestions should refer to meeting these goals more successfully or, alternatively, changing the goals to be more practical or relevant. As program goals often are broadly stated, it is incumbent upon the researcher to define these goals in a manner that is meaningful to program management. Many states have specific program goals as part of enacting legislation for drug courts, and it may be important to review these larger goals as part of the evaluation project. Supplemental to this is a determination of whether the program is operating as designed, particularly since it is not unusual to find that often there is a discrepancy between how a program was implemented versus how it was *intended* to be implemented (Longshore et al., 2001).

2. **Target Population** – It often is difficult to specifically define the population of offenders that a drug court program serves, considering the eligibility requirements that may or may not relate to the suitability of the client. However, it is essential to the operation of drug court programs that they be able to concisely identify the population they hope to serve and determine the extent to which they are reaching that intended group. Drug court evaluators should examine drug court client intake in terms of the program's stated goals (court goals as well as legislative, if applicable), resource limitations, and the universe of those who could be eligible for the program. A common complaint among many drug court programs is the inability to stay at full operating capacity. This problem can be researched and suggestions made through

a thorough analysis of client intake procedures and target population goals.

3. **Substance Abuse Treatment** – One of the aspects of drug courts that separates them from nearly all other justice system interventions is substance abuse treatment overseen using judicial monitoring and enforced supervision. To address this issue, it is crucial that baseline measures of addiction be considered. Evaluators should compare treatment plans with the actual implementation by the court. When possible, it is also important to determine the appropriateness of specific treatment modalities for particular clients. While it is not the purpose of this paper to recommend specific screening instruments, it is important that drug courts document client use prior to the program to enable accurate comparison throughout the program; as such, any instrument used must contain measures of past and present prevalence and incidence of drug use, addiction severity, and drugs of choice. Screening and assessment instruments should contain measures of the appropriateness of particular modalities for particular clients (e.g., American Society for Addiction Medicine criteria) and must be both reliable and valid. Baseline data should then be compared to one or more reassessments of clients' addiction severity, both during and at the conclusion of the program.

4. **Court Processes** – All of the activities of the drug court program should be documented. Researchers should examine graduation, phase advancement, sanctions and incentives, supervision, and the various ramifications of drug testing, as well as the relationship between client needs and services rendered. Behavioral research supports the notion that the magnitude of the sanction or incentive should be proportionally consistent with the precipitating incident, so sanctions and incentives should be measured in relation to client behaviors (Skinner,

1950). Therefore, it is both possible and desirable to create a ratio of behaviors to sanctions or incentives with the goal of a one-to-one ratio.

A great deal of information has been published recently about behavioral controls of client behavior. For example, Douglas Marlowe and Kimberly Kirby published an article entitled "Effective Use of Sanctions in Drug Courts: Lessons from Behavioral Research" (1999). In this article, the authors describe the need for an overall individualized behavioral plan for clients based upon their personal histories and stations in life. Further, one of the most supported behavioral principles is the idea that certainty (i.e., the likelihood that an action, good or bad, will elicit a response) is perhaps the most important factor in creating client behavioral responses. Thus, measuring the relationship of client behaviors to programmatic responses is critical. Both the perceived magnitude of incentives or sanctions and the application schedule should be reviewed.

Other aspects of court process that bear mentioning are the supervision of clients and the coordination of court activities. Client supervision is one of the key components of the drug court model. Elements of supervision include client contacts and oversight of client activities (e.g., employer contacts). Further, the coordination of service application falls under the rubric of court processes. Questions should be asked about information sharing and team involvement in the decision-making process.

5. **Units of Service** – Drug court clients generally receive a variety of services while in the program. Each of these services should be documented in a manner that helps the program consider the benefits of particular services. A solid process evaluation will report if clients are gaining from particular programs or interventions. A unit of

service is a simple way of measuring and documenting all of the services provided by drug court programs. Included in this documentation should be medical and psychological services, job training and placement services, educational services, and any other service to which the client was linked by program staff. When considering units of service, it is important to document both the referrals and participation in the service provided. Benefits of services would most likely be assessed by asking clients in a consistent manner (i.e., customer satisfaction index) their feelings about the services provided. This attitudinal measurement strategy can provide a useful resource for management.

6. **Team Member Cooperation** – Drug courts are collaborative efforts. Their success or failure is dependent upon the constant "give and take" that replaces the traditional adversarial system. Some method of qualitative organizational research is useful to determine how well the drug court team functions as a unit. One simple method for collecting this type of information involves questioning team members individually as to their perceptions regarding the extent to which their input is considered when decisions are made by the drug court team.

7. **Community Support** – Community support is vital to program success. In some jurisdictions, the voting community selects team members, and courts often use local businesses to provide token incentives. There clearly is value to program management exploring the reactions—either positive or negative—to the drug court in the community it serves, as the court may eventually need local funding and support to survive. With this in mind, it is often valuable to assess the support of stakeholders and community leaders. This can be done using a survey or questionnaire asking specific questions

about their understanding of the model and its implementation.

Drug Court Planning Process

During the course of conducting the process evaluation, particularly the gathering of information through interviews with team members and other important stakeholders, an ancillary component should involve the review of the drug court's *planning* process. Understanding how and under what conditions the court came into existence can help inform an understanding of current processes and protocols, especially if a program has evolved since its inception in response to particular, and often unanticipated, circumstances or resource constraints. An understanding of this important component of the process evaluation can be gauged by asking questions such as:

- Were all appropriate key players brought in to serve as part of the drug court team (to help develop goals, objectives, policies and procedures, and the mission statement)? If not, who was missing? Was a representative from the mental health community at the table?
- Were enough team members assembled?
- Were adequate and appropriate planning trainings offered to all team members?
- Were all of the available community resources documented?
- Did all team members sign a release of information in order to share confidential information with each other? Were team members trained on confidentiality?
- How was the target population defined?

Beyond these seminal questions, drug courts must be considered as organisms that are growing and redefining themselves on a regular basis (Carey & Finigan, 2004).

There is often attrition among the ranks of drug court practitioners working in a particular program. Further, advents in treatment modalities and changes in behavioral approaches force programs to be somewhat flexible in their activities. And despite the fact that there is often natural incongruity over time in program operations, evaluators must consider these factors as part of the growth of the program rather than as separate incarnations. These factors require that a few additional questions be asked of the program:

- What programmatic or personnel changes have occurred over time?
- Are on-going training opportunities (including in-service and cross-trainings) provided to and utilized by all team members? Are these trainings worthwhile?
- What data collection system is being used? Or, how are records kept? Is the current system (computer or paper) working well for all team members?

WHAT DATA ARE NEEDED TO COMPLETE THESE EVALUATIONS?

[11] To answer the research questions mentioned above, there is a significant amount and type of data required. For this reason, trainings supported by BJA require that an evaluator be part of the initial planning team. The purpose of the involvement of an evaluator is to ensure that goals and objectives are measurable and meaningful, and to assist in the effort to collect the appropriate data. History suggests that this is one area in which drug court programs have failed to help themselves. As drug courts are local collaborations of disparate actors, often the programs rely on each of the partners to collect their own data in their own manner while important coordination and operational data is left behind. Indeed, NDCI's review of evaluation plans suggests that the single largest problem facing drug court evaluators is the lack of good data in useable form.

While a comprehensive list of data elements for drug court process evaluation and performance measurement is provided in the forthcoming research monograph (Heck, in press), there are some things that bear mentioning in this article. The first is that data must be collected on a consistent basis that provides for *reliability* and *validity*. Reliability means that the concepts and variables are measured consistently over time (Senese, 1997). This is a significant problem in drug courts, especially given the fact that programs have been buffeted with a variety of data collection scenarios over the course of their existence. Often, programs change midstream to adopt the newest model for data collection. While this may be in the best interest of the drug court for future data collection, it is the responsibility of the staff to ensure that existing data is not lost. This becomes a serious difficulty for researchers and evaluators as they attempt to track and document the historical activities of the court programs in relation to the outcomes.

Validity refers to the extent to which the data accurately reflect the operationalization of the concept or variable. There are four types of validity often considered when making judgments about data. *Face validity* is an assessment of the validity of the data based upon "what makes logical sense" (Senese, 1997). That is, is it reasonable to assume that the measures used accurately depict the construct being measured? For example, it makes some sense that arrest data would be a good measure of criminal activity and therefore a valid measure of recidivism. The second type of validity is *predictive validity*. Predictive validity refers to the extent to which the data accurately predicts the concept. For instance, appropriately applied sanctions and incentives lead to improved client behavior. Thus, a measure of the temporal proximity (celerity) between the action and the court response provides a good measure of the implementation of the behavioral model. Third, there is *content validity*. Content validity requires multiple measures of the same effect. Going back to our example of sanctions

and incentives, it would be important to consider swiftness, certainty, and appropriateness of the sanction or incentive to measure the implementation of the behavioral model. Insofar as these variables work together to predict the outcome, they exhibit a high degree of content validity. Finally, there is *construct validity*. Construct validity refers to the extent to which the measures reflect what is theoretically predicted by the research design. In drug court theory, it is assumed that the confluence of the behavioral, supervisory, and treatment processes lead to client success. Thus, measuring the behaviorism (i.e., incentives and sanctions) in the program process would have construct validity as part of the overall model (Senese, 1997).

Thus, generally speaking, the data elements that are collected by drug court programs must be both reliable and valid. The second important point is that drug court data collection must also capture the important variable of time. Time is generally captured by date stamping all drug court activities. For example, it is important to collect information regarding the date in which a drug court infraction occurred as well as the date when the sanction was applied. These dates allow evaluators to measure the time gap to effectively consider the issue of celerity. It is impossible to measure client performance and improvement without documenting the dates of all activities. This should perhaps be understood intuitively, but unfortunately there are many examples of data collection efforts, particularly in areas of drug testing and incentives and sanctions, which forget this important component.

The third major point is that the best time to start collecting this data in a uniform manner is now. The lack of valid and reliable data from which to assess drug court program performance creates a series of issues for evaluators and researchers. The best way to address these potential problems is to avoid them in the first place. This can be done by carefully conceptualizing the model using the available

resources and working with an evaluator to ensure that the correct data are being collected in a usable format from the inception of the program. If the program is already operational it is strongly recommended that these data concerns be addressed as soon as possible.

Several states and localities are developing or have developed standardized and comprehensive electronic systems for capturing the important data. It is highly recommended that these systems serve as more than just data repositories. There are myriad examples of social programs that have developed data collection systems that have been poorly managed and provide little service to the local programs. Indeed, it seems that this lesson is generally hard to learn, as data collection systems for drug courts are still being developed even at the national level. The technology exists to create case management systems that provide the users with utility and promote the input of valid data for the purposes of managing local and state programs. As more drug court specific systems are developed, the costs are shrinking and the benefits of such a system are tremendous both for program evaluation and in a broader sense, sustainability.

Each of the basic elements mentioned above have attendant variables associated with them. Using the concepts of reliability and validity as a guide, evaluators, program managers, and stakeholders need to develop a data collection enterprise that accurately reflects the program components based upon the theories guiding the model. Again, expediency is critical. It is not enough to allow the partners in the program to collect their own required data, as the important elements of cross-pollination will be missed and the likelihood of identifying gaps in services will be reduced.

It is also important to remember that drug court programs involve treatment providers that collect confidential information that is subject to the Health Insurance Portability

and Accountability Act (HIPAA, 1996) and federal medical confidentiality regulations (CFR 42). These standards require careful management of data and must be followed to ensure continued licensure. There is some confusion about requirements related to sharing health care information. Many people incorrectly assume that HIPAA disallows the sharing of any health care information, but this is not strictly accurate. HIPAA does not limit the ability of direct care providers to share information as long as it is for the benefit of their patients' treatment. There are, however, substantial requirements related to informing and gaining consent from patients (U.S. Department of Health and Human Services, 2003). Evaluators must work with program management to develop the appropriate waivers for clients and to make sure that confidentiality lines are not crossed.

WHAT DOES IT MEAN TO BE "METHODOLOGICALLY RIGOROUS?"

[12] Methodological rigor refers to using empirical and scientific models for analyzing cause and effect. While strong in developing correlations and relationships, social science as a whole has difficulty establishing true causality using the scientific method, and drug court research is no exception. While there are ongoing studies that use random sampling and empirical design (see Marlowe et al., 2004), most local courts and evaluators do not have the available resources to conduct such research. Indeed, by design, evaluation research is not meant to create high statistical significance and be able to claim true scientific causality.

That is not to say, however, that evaluations should not be rigorous in following accepted protocols and methodology. In fact, one of the major problems with many current evaluation endeavors is that they do not follow any uniform protocols. Professional evaluators have developed general standards which are appropriate to apply to drug court evaluation. These standards include *Utility Standards*,

designed to ensure that user's needs will be met; *Feasibility Standards*, designed to ensure that the evaluation will be "realistic, prudent, diplomatic, and frugal"; *Propriety Standards*, designed to ensure that evaluations will be legal and ethical; and *Accuracy Standards*, designed to ensure that the measurement obtained matches the actual value of the variable being measured (Sanders & Joint Committee for Standards for Educational Evaluation, 1994). These standards can be referenced and used as tools for guiding contracts with evaluators as well as for assuring the quality of the evaluation product. While they do not necessarily guarantee academic rigor, they do serve as a good starting point for dialogue with potential evaluators.

Additionally, there are several methodological issues in evaluation that can be avoided by using researchers who understand the foundations of scientific inquiry. It is often the case that evaluators try to do too much with the limited resources available to them. While this effort is somewhat laudable, the result can be poorly conceived research, which ultimately leads to a continued lack of credibility for drug court research as a whole. One common problem relates to attempts to claim the use of experimental design in local program evaluation when there clearly is no true experimentation involved.

WHAT ABOUT EXPERIMENTAL DESIGN AND COMPARISON GROUPS?

[13] In order to assess the relative impacts of drug courts, it is necessary to compare the program effects to those of other similarly situated offenders. However, this comparison is not unilaterally required for process evaluation. Process evaluations can serve the purposes of documenting program development, assessing the extent to which goals and objectives are being met, and ensuring fidelity to the model, without comparison groups. Many times this is enough for program managers and stakeholders. Indeed,

doing this much in a thorough and useful manner is preferable to ill-advised attempts to create a comparison.

However, comparison is the only way to document the impact of the program on clients relative to other interventions. True experimental design is the gold standard for this type of research. Using random assignment of subjects and controlling for extraneous pressures provides researchers with the best method for claiming causality (Campbell & Stanley, 1963). Random subject assignment allows researchers to argue that the groups in the study are equal. This "equivalence" is central to making claims about the true effects of the intervention. Unfortunately, experimental design is extremely difficult to perform in a real-world, criminal justice setting and often inspires questions concerning ethics and fairness. These concerns generally focus on issues of equal access when considering designs that exclude groups from accessing drug court services. Additionally, local programs often do not have the resources (both financial and in the number of available clients) to perform such studies.

The next best approach for documenting the effects of programs is undertaking a quasi-experimental approach using *comparison groups*. Comparison groups do not afford the researcher the ability to claim "equivalence" between the treatment group and the non-treatment group, however, they do, in more general terms, provide evaluators with some measure of program effect. Comparison groups are only as good as the specificity upon which they are being compared. For example, comparing drug court clients to all offenders in a state is a weak comparison. And, while weak comparisons can sometimes be better than no comparison at all, there are many factors that make these two groups systematically dissimilar and thus render claims of relative program success quite tenuous.

In developing comparison groups, researchers should first identify the primary and secondary factors that are related to outcomes as suggested in the literature and match their comparison groups on these items. For drug courts, these factors include some basic demographics (e.g., age and gender) as well as research-driven individual factors (i.e., anti-social personality disorders, criminal history, and previous treatment failures). Some of these variables are continuous (they can be scaled), while others are dichotomous. Thus, it is important to match continuous with continuous and dichotomous with dichotomous. Often, the availability of data determines the extent to which comparisons can be made. Great care must be taken in the selection of these groups (Maxfield & Babbie, 2005).

A common mistake made by drug court programs has been to use dropouts and program failures for comparison. There are at least two reasons why this is inadvisable. First and foremost, there are clear systematic differences between program completers (or graduates) and those who leave the program before completion; these differences are impossible to disaggregate during analysis. Many hypotheses exist about why some individuals are more likely than others to complete drug court programs. Almost all of these hypotheses are untested and unproven and as such, statistical control for these variables sheds little light on the variation. Second, while many program failures and dropouts leave in the early stages of the program, it is impossible to identify the relative effects of limited exposure to the program and thus the analysis will be clouded by uncertainty. Clients who stay in the program for any period of time are exposed to the model and thus are tainted for true comparison.

CONCLUSION

The eminent philosopher Thomas Kuhn suggests that science is designed to solve puzzles (1970). Program evaluation is an important tool for program management and

improvement, as well as for solving puzzles such as what makes these drug court programs work. It provides a means by which drug courts can document progress, memorialize actions, and maintain accountability as well as the fidelity of the model. While there have been many excellent program evaluations conducted on drug court programs, there remains considerable confusion and inconsistency surrounding this important endeavor. It is hoped that this article will help to create a foundation upon which drug courts can build an evaluation model that leads the field of criminal justice and supports continued growth and improvement.

REFERENCES

Bachman, R., & Schutt, R.K. (2003). *The practice of research in criminology and criminal justice* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.

Carey, S.M. & Finigan, M.W. (2004). *Adult drug court typology interview guide*. Portland, OR: NPC Research, Inc.

Code of Federal Regulations Title 42.

Goldkamp, J. S., White, M.D. & Robinson, J.B. (2001). Do drug courts work? Getting inside the drug court black box. *Journal of Drug Issues 31*(1), 27-72.

Heck, C. (in press). *Local drug court research: Navigating performance measures and process evaluations.* Alexandria, VA: National Drug Court Institute, National Association of Drug Court Professionals.

Health Insurance Portability and Accountability Act (HIPAA). (1996, August 21). Public Law 104-191. 104th Congress.

Huddleston, C.W., Freeman-Wilson, K., Marlowe, D.B., & Roussell, A.P. (2005, May). *Painting the Current Picture: A National Report Card on Drug Courts and Other Problem Solving Courts, I*(2). Alexandria, VA: National Drug Court Institute, National Association of Drug Court Professionals.

Kuhn, T.S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Longshore, D., Turner, S., Wenzel, S., Morral, A., Harrell, A., McBride, D., Deschenes, E., & Iguchi, M. (2001). Drug courts: A conceptual framework. *Journal of Drug Issues 31*(1), 7-26.

Marlowe, D.B., DeMatteo, D.S., & Festinger, D.S. (2003, October). A sober assessment of drug courts. *Federal Sentencing Reporter, 16*(1), 113-128.

Marlowe, D.B., Festinger, D.S., & Lee, P.A. (2004). The judge is a key component of drug court. *Drug Court Review IV*(2), 1-34.

Marlowe, D.B. & Kirby, K.C. (1999). Effective use of sanctions in drug courts: Lessons from behavior research *Drug Court Review II*(1), 1-31.

National Association of Drug Court Professionals. (1997, January). Defining drug courts: The key components. Washington, DC: Bureau of Justice Assistance, U.S. Department of Justice.

Office of Justice Programs. (1998). *Drug court monitoring, evaluation, and management information systems*. Washington, DC: Author, U.S. Department of Justice.

Rossi, P. H. & Freeman, H.E. (1993). *Evaluation: A systematic approach* (5th ed.). Thousand Oaks, CA: Sage Publications.

Sanders, J., & The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Senese, J.D. (1997). *Applied research methods in criminal justice*. Chicago: Nelson Hall Publishing.

Skinner, B.F. (1950). *Science and human behavior*. New York: McMillan Publishing.

U.S. Department of Health and Human Services. (2003). *Fact sheet: Protecting the privacy of patients' health information* [Online]. Available: www.hhs.gov/news/facts/privacy.html.

U.S. Government Accountability Office. (2005, February). *Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes*. Report to congressional committees. Washington, DC: Author.

RECIDIVISM 101: EVALUATING THE IMPACT OF YOUR DRUG COURT

By Michael Rempel
Center for Court Innovation

Since their inception, drug courts have consistently sought to reduce the recidivism rates of their participants. Despite the centrality of this goal, drug court administrators, staff, and local evaluators often have questions about how to conduct a valid recidivism analysis. This article provides an accessible introduction to the following key methodological issues: (1) how to define recidivism (e.g., re-arrest, re-conviction, or re-incarceration), (2) which drug court participants to include in the analysis (all participants or a select sub-sample), (3) how to construct an appropriate "comparison group" (composed of defendants who did not enroll in the drug court but who are likely to be similar in their characteristics), and (4) how to ensure statistically that the final drug court and comparison samples are in fact highly similar (same distribution of key socio-demographic measures, criminal history, and current charges). If practitioners and evaluators alike develop a basic comprehension of the key methodological issues, they can become productive partners in the implementation of any recidivism research.

This research paper was originally supported by the Bureau of Justice Assistance, Office of Justice Programs, U.S. Department of Justice (contract #98-DC-VX-K007). Any opinions and interpretations expressed are those of the author and do not necessarily represent the official position of the U.S. Department of Justice. The author would like to thank Donald J. Farole, Jr., Greg Berman, Valerie Raine, Robert V. Wolf, and three anonymous reviewers for the Drug Court Review *for their valuable feedback and guidance.*

Michael Rempel is research director at the Center for Court Innovation in New York City. His research focuses primarily on drug courts, domestic violence courts, and other "problem-solving justice" initiatives. He was recently principal investigator of a statewide evaluation of adult drug courts in New York and is currently co-principal investigator of a national drug court evaluation conducted with the Urban Institute and the Research Triangle Institute. He has also participated in several research projects concerning the potential to apply the "problem solving court" approach more broadly throughout state court systems.

Direct all correspondence to Michael Rempel, Center for Court Innovation, 520 8th Avenue, 18th Floor, New York, New York 10018. (212) 373-1681; (212) 397-0985 (fax); rempelm@courtinnovation.org.

ARTICLE SUMMARIES

WHAT DO WE KNOW NOW?

[14] It has been established that drug courts work; differences in recidivism rates are often products of different geographical areas, styles of court, and different target populations.

WHAT IS "RECIDIVISM"?

[15] Recidivism is usually defined as rearrest, although other data is valuable. Timeframe of analysis is an important factor, but is often limited by the age of the drug court.

WHICH DRUG COURT PARTICIPANTS SHOULD BE INCLUDED IN THE ANALYSIS?

[16] A representative sample of participants should be included, inclusive of graduates and failures.

WHAT IS AN APPROPRIATE COMPARISON GROUP?

[17] There are numerous options for comparison groups depending on the situation of the drug court, but some are clearly superior to others.

HOW DO YOU ENSURE THAT THE DRUG COURT AND COMPARISON GROUP SAMPLES ARE TRULY COMPARABLE?

[18] It is crucial to determine the degree of similarity between the comparison group and drug court group. A trained evaluator has a number of statistical methods available, of which the drug court staff should have a working knowledge.

INTRODUCTION

Amidst widespread agreement that producing reductions in recidivism is an important goal of the criminal justice system and a universal goal of drug courts in particular, drug court administrators and staff routinely query how to go about conducting a valid recidivism analysis. While trained evaluators usually do the work, their range of expertise and possible methods from which to select are considerable. If drug court staff themselves had a basic understanding of the key methodological issues, they could become more active partners in the research design and analysis. This would help both the evaluators by providing a new source of informed feedback and drug court staff by increasing their trust and comprehension of the ensuing results.

In an attempt to build a bridge between practitioners and evaluators, this paper provides an overview of four methodological questions that must be addressed in any recidivism analysis. Examples from the evaluation literature are incorporated throughout to show how different methods have been applied. For overview purposes, the four questions are:

1. *What is "recidivism"?* What recidivism measures are appropriate: re-arrest, reconviction, re-incarceration, or others? What is the ideal timeframe for measuring recidivism: one year after drug court participation begins, two years after participation begins, or one year after program exit?

2. *Which drug court participants should be included in the analysis?* Is there a generally accepted definition of the universe of "drug court participants" to be considered in any recidivism analysis? Should recidivism rates be computed for all participants who have ever entered the program, or are there good reasons to exclude certain categories?

3. *What is an appropriate comparison group?* What *is* a comparison group? What are the most popular comparison group designs and their respective advantages and shortcomings?

4. *How do you ensure that the final drug court and comparison group samples are truly comparable?* Having established what *seems* like appropriate drug court participant and comparison group samples, is it possible to verify whether they are truly comparable? If they are not—if they differ in demographics, charges, criminal histories, substance abuse histories, or other important background characteristics—attempts to compare their recidivism rates could produce biased results. If potential biases are found, are there methods for correcting them?

Before turning to these questions, the next section reviews what we already know about the impact of drug courts on recidivism. This serves to establish realistic expectations for interpreting future results.

WHAT DO WE KNOW NOW?

[14] Drug courts usually reduce recidivism. Most studies report *lower* recidivism rates among drug court participants (including both graduates and failures) than similar defendants prosecuted in a conventional fashion. In one recent review of the literature, David Wilson and colleagues found that the recidivism rate, defined in most studies as the re-arrest rate, was lower among drug court participants than among other similar defendants in 37 of 42 sites evaluated, and was lower by an average of approximately 13 percentage points (e.g., from 50 percent to 37 percent), with some programs producing much larger and some much smaller effects (Wilson, Mitchell, & MacKenzie, 2002).

While this review is extremely positive, much of the recidivism literature, particularly the first generation of studies completed in the 1990s, possesses serious methodological shortcomings. Most notable is a failure to identify an appropriate "comparison group" of defendants with whom drug court participant outcomes could be reasonably compared (see critiques in Belenko, 2001; Roman & DeStefano, 2004). For example, as will be discussed below, studies comparing recidivism between drug court graduates and failures, or comparing drug court participants to those found ineligible for the program, are not valid. Fortunately, most researchers would agree that the quality of the evaluations produced in the early 2000s greatly improved on the earlier efforts. Consequently, three other recent reviews which considered a smaller number of drug court evaluations, mainly by eliminating ones with weak methodologies, still reported lower recidivism rates among drug court participants than comparison group defendants in nearly all sites examined (see Aos, Phipps, Barnoski, & Lieb, 2001; U.S. Government Accountability Office, 2005; Roman & DeStefano, 2004).

Most of the evaluations included in these reviews examined re-arrest rates over a one- or two-year period after the initial arrest that led either to drug court participation or to inclusion in the comparison group. While only a handful of evaluations have isolated *post-program* recidivism (after participants have either graduated or failed), their results are also encouraging. A study of six New York State drug courts reported consistent recidivism reductions over a one-year post-program period—an average 31 percent reduction relative to the comparison group level during a comparable one-year period (Rempel et al., 2003). A study of the Los Angeles County drug courts similarly isolated recidivism during a one-year post-program period (Fielding, Tye, Ogawa, Imam, & Long, 2002). Interestingly, this study found that the drug court produced significant recidivism reductions among "medium" and "high" risk defendants but

not among "low" risk defendants; risk level was defined by a combination of defendant prior criminal history, severity of the current arrest charges, and community ties (e.g., employment status and living situation). Several other studies have confirmed that various aspects of the drug court model work particularly well with high risk defendants (see Marlowe, Festinger, & Lee, 2004; Rempel & DeStefano, 2001).

Although the research literature is clear that not all drug courts produce effects of the same magnitude, the available evidence demonstrates, overall, that the model works. Thus in a recent review, Douglas Marlowe and his colleagues concluded, "The best available research evidence suggests that drug courts can reduce drug use and criminal recidivism on an order of magnitude of two to three times greater than almost any other initiative that has been attempted with this intransigent population" (Marlowe, DeMatteo, & Festinger, 2003, 153). At the same time, most drug courts do not achieve the monumental effects sometimes claimed by overly enthusiastic proponents, often creating an unfortunate expectations gap. To wit, few drug courts cut the recidivism rate by as much as half; in fact, reducing recidivism by as much as a quarter relative to baseline levels (e.g., reducing the re-arrest rate from 40 percent to 30 percent) is a respectable and commendable achievement for any criminal justice intervention. By setting realistic targets, drug courts can position themselves to conduct well-designed evaluations and learn from their results without facing political pressures to attain the unattainable.

FOUR QUESTIONS CONCERNING RECIDIVISM METHODOLOGY

What Is "Recidivism"?

[15] Most completed drug court evaluations define recidivism as re-arrests; some also use reconvictions instead

or in addition (see studies reviewed in Wilson et al., 2002; or Government Accountability Office, 2005). Arrest-based measures are often preferred for several reasons. First, sometimes cases may be dismissed or pled down to levels falling short of a criminal conviction for technical, evidence collection, or criminal-history related reasons that may not reflect the absence of criminal behavior. Second, since arrests usually follow shortly after the underlying criminal behavior takes place, the use of re-arrest measures makes the timeframes for analysis fairly straightforward. By comparison, months may pass between a re-arrest and a reconviction, and these case-processing delays may complicate the analysis. For example, a one-year recidivism analysis using re-convictions may, in practice, require that the underlying criminal behavior take place within a much shorter timeframe to allow extra time for both the criminal behavior and the dispositional process to be completed within the allotted year. Nonetheless, persons are not always guilty as charged and thus an analysis based on reconvictions retains the advantage of filtering out weak cases or ones where innocence may subsequently have been established.

In drug courts that only accept defendants arrested on drug charges, it may also be advantageous to isolate recidivism on drug-related charges. Breakdowns for felony as opposed to misdemeanor recidivism may be revealing as well. For example, in a study of the Escambia County, Florida drug court, the researchers found that there was not a significant difference between drug court participants and the comparison group in re-arrest rates for *all* types of offenses, but when isolating results for more serious *felony* offenses, the re-arrest rate was significantly lower for drug court participants (Truitt, Rhodes, Seeherman, Carrigan, & Finn, 2000).

Finally, if one is particularly interested in cost savings issues, it may be advantageous to look at measures of re-incarceration. If drug court participation leads to a

significantly lower prevalence of new crimes that result in lengthy jail or prison sentences, then the drug court may be able to achieve meaningful cost savings for local and state correctional agencies.

As important as the choice of measure (re-arrest, reconviction, or re-incarceration) is arguably the choice of timeframe. Most studies have defined their timeframe to begin at the outset of drug court participation (and at an equivalent early date for the comparison group). This means that recidivism is mainly considered during an *in-program* period of time and, when the measurement period extends for two years or longer, for perhaps a little bit of *post-program* time as well. Evaluating recidivism in this way—largely during an in-program period—is important, because it tests whether judicial supervision by the drug court can produce an immediate impact in suppressing criminal behavior. However, drug courts often present themselves as having long-term behavioral effects. Therefore, evaluating post-program recidivism, after drug court graduation or failure, provides a critical measuring rod of whether drug courts have really achieved all of their goals (see discussion in Belenko, 2001).

Post-program analyses, however, have an important practical disadvantage: It may take years for enough participants to enroll, graduate or fail, and then accumulate a sufficient amount of post-program time in the community for a post-program recidivism analysis to begin. Therefore, drug court staff that would like to see some recidivism results on a more timely schedule, should argue for foregoing, or at least postponing, a post-program analysis in favor of an in-program one.

Whatever timeframe is selected should be identical or at least equivalent for both drug court participants and the comparison group. Also, if a post-program timeframe is selected, particularly for drug court failures and for the

comparison group, it is important to begin the time count not on the drug court failure or final disposition date, but on the date of release from jail or prison in the event that the defendant was incarcerated. To understand why this is important, consider the case of a defendant who fails drug court and is sentenced to one year in a jail: a one-year post-program analysis would presumably find that such a defendant did not re-offend for the simple reason that the defendant was serving a jail sentence during the entire one year measurement period and hence not "at risk" of re-offending in the community. For such a defendant, it is therefore necessary to measure recidivism over the second year that begins *after* the initial year spent in jail.

Conclusion: *Recidivism is usually defined as re-arrest but sometimes as reconviction or re-incarceration; in evaluating some drug courts, it may also make sense to isolate recidivism on certain kinds of offenses (e.g., felony, misdemeanor, or drug-related offenses). The drug court staff should feel free to discuss with the evaluator its own preferences for defining recidivism. Furthermore, staff should express its preference for the analysis timeframe, recognizing the tradeoff—a longer timeframe (e.g., one year post-program) will enable testing the long-term behavioral effects of the drug court, but a shorter timeframe (e.g., one or two years post-intake) will enable conducting the analysis and providing results after fewer years have elapsed.*

Which Drug Court Participants Should Be Included in the Analysis?

[16] The next step in designing a recidivism analysis is to determine the universe of "drug court participants." Ideally, it should consist of a representative sample of all participants and should be large enough to produce results that cannot be attributed to chance.

In most drug courts, identifying the participant pool is straightforward, since all of them must sign a contract upon enrolling or, in many cases, plead guilty to some offense. Once participation is formalized, the person qualifies for a recidivism analysis. This is the case even if the person disappears from program contact the very next day, never to be seen again.

In this regard, it cannot be emphasized enough that "participants" means all participants, not merely successful ones. To address a common misunderstanding, it is invalid to highlight the performance of graduates alone in attempting to determine whether a drug court reduces recidivism. It may be informative to know the performance of the graduates; for example, if the recidivism rate for graduates is very low, one response might be to implement revised policies or additional services designed to increase the graduation rate. Nonetheless, recidivism results for graduates by themselves do not have *evaluative* significance. As a policy matter, what is important to know for impact evaluation purposes is how the drug court fared with everyone it attempted to serve: Does a policy of routing defendants to drug court produce better outcomes for the system than not doing so? The answer obviously depends on what happens to everyone so routed. No one would consider a program successful if only 10 percent of its participants graduated, even if that 10 percent had a miniscule recidivism rate. Further, even if it appears that drug court graduates are performing particularly well, it cannot be inferred that the drug court was the cause; perhaps those defendants that graduated had already grown tired of their former lifestyle and would have avoided re-offending in any case, with or without the drug court intervention.

Does this mean that it is necessary to include every participant in a recidivism analysis? Not necessarily. First, it may be desirable to exclude those enrolling at the outset of the program, when the drug court may have been building up to capacity, initiating policy refinements, still implementing

data collection systems not yet in use, or working out other kinks in its operations. For instance, in evaluating the Rochester (New York) Drug Treatment Court, the researchers decided to exclude drug court participants enrolling in 1995, the first year of operations, because it was the first to open in New York State and had to develop much of its model after operations began (Rempel et al., 2003). Another approach may be to include participants enrolling in all years, but to conduct separate analyses for different years of entry, so that changes in effectiveness over time can be captured. Also, one generally excludes from a recidivism analysis the most recent drug court entrants, since they will not have been in the program for long enough to have their recidivism rates tracked. For this reason, recidivism analyses are difficult to conduct soon after a drug court opens. It is necessary to wait, sometimes for years, until enough participants have accumulated enough time after program entry to qualify them for an analysis spanning a meaningful timeframe (at least one year post-entry and preferably longer). Finally, when attempting to analyze recidivism over a *post-program* period

Case-in-Point: The Portland, OR and Las Vegas, NV Drug Court Evaluations*:* In a two-site study of the Portland and Las Vegas drug courts, John Goldkamp and colleagues (2001) addressed the implementation issue that drug court performance can change over time by conducting separate recidivism analyses for each year's cohort of drug court participants and comparison group members. In both sites, the evaluation found that the magnitude of the drug court's effects on recidivism varied substantially by year of entry. Those entering the drug court in some years had substantially lower recidivism rates than that year's comparison group, whereas those entering in other years did not fare differently than the comparison group. The authors attribute these results to changes in the Portland and Las Vegas drug court programs over time, leading the programs to be more efficiently and effectively run in some years than others.

after graduation or failure, it goes without saying that only graduates and failures should be included, not participants who are still actively engaged in the drug court program.

A separate consideration is the *number* of available participants. The general rule is that the greater the sample size, the smaller the margin of error for each reported recidivism rate, although adding more sample size is far more helpful at the low end of the spectrum (e.g., going from 50 to 100 participants) than at the high end (e.g., going from 400 to 500 participants).

What precise sample size is sufficient for a given recidivism analysis? In most cases, a sample size of at least 100 participants and possibly more is necessary to generate "statistically significant" results that fall outside the study's margin of error. To clarify precisely how large a sample size is required, researchers will commonly use a method called "power analysis." Such an analysis helps to project how large a sample is necessary to determine if two populations (e.g., drug court participants and a comparison group) have a "statistically significant" difference. To illustrate, in the table below, we assume that the comparison group has a re-arrest rate of 50 percent and, for several different sample sizes, conduct a power analysis to determine what the drug court re-arrest rate would have to be for the difference to reach statistical significance. With just 50 participants and 50 comparison group defendants, the drug court recidivism rate would have to drop from 50 percent to 22 percent or less to achieve significance. A difference of this magnitude would be close to unprecedented in the drug court literature. Although some drug courts have been able to achieve the impact that would be required with sample sizes of 100 (e.g., a reduction from 50 percent to 30 percent in the re-arrest rate), most drug courts have fallen short of this magnitude as well. Therefore, it is only as the samples grow much larger than 100 does it become likely for the average successful program actually to show a statistically significant effect.

With samples of 200, the drug court need only show a reduction in the re-arrest rate from 50 percent to 36 percent to reach significance, a magnitude that approximately half of all drug courts studied to date *have* achieved. Interestingly, once the sample sizes grow extremely large, further additions do not take on as much importance. For instance, as shown below, little is gained from increasing the sample sizes from 600 to 800.

Table 1. Sample Sizes and Corresponding Arrest Rate Significance

Comparison Group Sample Size	Drug Court Sample Size	Comparison Group Re-Arrest Rate	Drug Court Re-Arrest Rate Needed to Achieve Significance
50	50	50%	22% or less
100	100	50%	30% or less
200	200	50%	36% or less
400	400	50%	40% or less
600	600	50%	42% or less
800	800	50%	43% or less

This discussion suggests that large sample sizes are essential to generate statistically meaningful results. Yet, it is important to keep in mind that many drug courts are inherently constrained by serving only a small volume of participants. In a sense, it is therefore impractical to require all drug court evaluations to achieve the kinds of sample sizes that are ideal from a pure statistical perspective. Furthermore, limiting the evaluation literature to drug courts able to generate large samples may prevent the field from gaining information about the operations and effects of smaller programs that are located in more rural settings. In this light, local evaluations can and probably should still proceed even with small samples. As long as the statistical limitations to such evaluations are plainly understood and acknowledged, the results can perhaps be suggestive in

themselves and informative to practitioners and researchers planning future evaluations with larger samples.

Conclusion: *A representative sample of drug court participants (including both graduates and failures) should be included. Since sample size is a critical factor affecting the potential for a recidivism analysis to produce statistically significant results, drug court staff should communicate its rate of intake to the evaluator early on and help the evaluator to develop a realistic timeline for accumulating a sufficient sample to conduct the analysis.*

What Is an Appropriate Comparison Group?

[17] The performance of drug court participants becomes meaningful only in relation to a "comparison group," defined as a group of defendants who did not enter drug court but are similar in their criminal justice status and other characteristics (e.g., demographics, substance abuse history, criminal history). It is important for the background characteristics of the comparison group to be as similar as possible to the participants; otherwise, the recidivism results may be misleading. To illustrate why this is so, consider the implications of having dissimilar samples with respect to prior criminal history. It is well known in criminology that defendants with more prior offenses are more likely to commit future offenses. Therefore, if the drug court participant sample averages fewer priors than the comparison group, and if participants have a lower recidivism rate, this difference in recidivism may be attributable merely to the participant sample's overall reduced criminal propensity, not to the positive impact of the drug court intervention *per se*.

The following provides a brief survey of popular comparison group designs in approximate order of quality (highest to lowest).

Randomized Trial. This is the "gold standard." First, defendants are screened to determine whether they are eligible for the drug court. Those who are eligible and willing to participate are then *randomly assigned* to either the drug court or the comparison group. In theory, the random assignment process ensures that defendants in both samples will be nearly identical in all ways besides their drug court participation status. This is because the only difference is the "luck of the draw" at the time of randomization. In practice, these designs are often impractical. They raise the ethical dilemma of denying a treatment thought to be effective to the comparison group and can raise the implementation problem of requiring a program to operate under capacity, since the random assignment process will re-route roughly half of the eligible pool to the comparison group. Also, randomized trials are not always unassailable methodologically. The research integrity of such trials may be compromised if judges or other court staff can selectively remove large numbers of defendants from the randomization process; or if the drug court changes its eligibility criteria during the period of the study, for example by allowing only defendants arrested on less serious charges to participate in the random assignment. Nonetheless, generally well-implemented randomized trials have been conducted on three adult drug courts in Washington, D.C., Maricopa County, Arizona, and Baltimore, Maryland, and of a juvenile drug court in Summit County, Ohio.

Case-in-Point: The Baltimore City Drug Treatment Court Evaluation*:* This is one of the most highly-regarded drug court evaluations in the literature. It involved the random assignment of 235 defendants to either (1) participation in the Baltimore City Drug Treatment Court or (2) conventional case processing. The random assignment took place after a defendant was determined to be eligible for the drug court. After random assignment, as in many studies of this nature, the judge or other key officials could opt at their discretion to remove individuals from their randomly assigned condition; however, in this particular study, officials altered the random assignment of only 9 percent of those who had been assigned to the drug court and only 7 percent of those who had been assigned to conventional case processing. These are extremely low change rates relative to other random assignment studies in the literature, suggesting a well-implemented research design. Two years later, 66 percent of those assigned to the drug court and 81 percent of those assigned to conventional case processing were re-arrested; three years later, the respective re-arrest rates were 78 percent and 88 percent, again with those assigned to the drug court re-arrested at the lower rate (Gottfredson, Najaka, & Kearley, 2003; Gottfredson, Najaka, Kearley, & Rocha, 2003).

Contemporaneous and Not Screened for Drug Court. In general, a "contemporaneous" comparison group includes defendants who did not enroll in the drug court even though they were arrested during the tenure of the drug court. In assessing contemporaneous designs, the first question is *why* the potential comparison group members did not enroll: Did the prosecutor oppose their participation? Were they found not to be drug-addicted? Did they refuse to participate? Or did other factors lead them to be ineligible? For example, defendants not entering the drug court due to a

refusal to participate may start with less motivation to change their behavior and may therefore be inherently more likely to re-offend in the future.

As a general rule, the best contemporaneous designs involve defendants who were formally eligible for the drug court but were never screened for it for strictly logistic, bureaucratic, or organizational reasons. For example, if a drug court caps its caseload at a certain level, those not participating strictly for lack of program capacity would comprise a good contemporaneous comparison group. Or if bureaucratic mistakes lead some defendants not to be referred to the drug court when they should have been, such a development could also make for a good comparison group. One disadvantage of these kinds of comparison groups is that since the defendants may never have been assessed by drug court staff (e.g., because they were never referred in the first place), it is usually unknown whether or not they are addicted to drugs. Instead, their comparability to drug court participants is often based on more formal criteria such as their criminal history, current charges, or basic demographics that may be obtainable from court records.

Case-in-Point: The Rochester Drug Court Evaluation*:* In a recent evaluation of the Rochester Drug Treatment Court, the researchers took advantage of a lack of political support for the drug court and consequent unwillingness to refer cases among all but two judges on the arraignment circuit (Rempel et al., 2003). The comparison group consisted of defendants arraigned on drug court-eligible charges by a judge other than those two. The evaluation showed a small but significant drug court impact over a one-year *post-program* period—the reconviction rate was 42 percent for drug court participants compared with 48 percent for the comparison group.

Pre-Post. A "pre-post" design compares drug court participants to similar defendants arrested before the drug court opened, often in the year prior. Again, with this design, it may not be possible to obtain data on whether comparison group defendants are drug-addicted. Instead, the comparability of participants to comparison group defendants may be based solely on data that is obtainable from official court records. Also, unlike a contemporaneous design, a "pre" comparison group may be vulnerable to what is known as "historical bias." This kind of bias arises if police deployment patterns, prosecutorial strategy, or relevant local laws significantly changed before and after implementation of the drug court. Those changes may have affected the natural probability that defendants in the "pre" as opposed to the "post" samples will be re-arrested for the same behaviors. For instance, after September 11, 2001, some police officers in New York City were re-deployed from investigating narcotics crimes to engaging in counter-terrorism efforts, thereby reducing the prevalence of drug arrests during the immediate post-9/11 period. In general, however, police and

Case-in-Point: The Bronx Drug Court Evaluation*:* The Bronx was one of the additional sites involved in the statewide evaluation of New York's drug courts. Unlike Brooklyn and Rochester, a strong contemporaneous design was not feasible, so a pre-post design was used instead. The Bronx supported a particularly strong pre-post design: As a result of the high volume of drug court-eligible defendants in the county, the entire comparison group was obtainable from the pool of defendants arrested during only a four-month period immediately preceding the outset of drug court operations. This made the chances of "historical bias" extremely small. The analysis found that over a one-year post-program period, the reconviction rate for Bronx Treatment Court participants was 16 percent, compared with 29 percent for the comparison group (Rempel et al., 2003).

prosecutors are not constantly changing their practices, so the mere potential for historic bias should not deter a drug court from exploring the pre-post design option if the choice is available.

Refused Treatment. A "refused treatment" comparison group includes those screened and found eligible for the drug court but who refused to participate. As noted above, this design may have a critical shortcoming—refusers may lack interest or motivation to participate, making their baseline situation fundamentally different from real participants. Refused treatment comparison groups are nonetheless extremely popular, since they are often easy to obtain—many drug courts record when a defendant is screened but refuses to participate. While refused treatment comparison groups are therefore to be viewed with caution, they are not equally problematic in *all* cases. In some situations, the reasons why some defendants refuse to participate may not necessarily create an obvious bias. For example, if certain defense attorneys in a jurisdiction advise their clients to refuse, while other defense attorneys do not, then the characteristics of actual defendants may not really differ between participants and refusers. For this reason, such an approach should not be dismissed outright. A helpful first step would be for drug court staff to lead the evaluator through the drug court's screening process, so that the reasons why some defendants opt not to participate can be better understood. Then an informed decision can be made about the likelihood and degree of bias that would be introduced by a refused treatment approach.

Comparison Jurisdiction. This type of comparison group consists of defendants who meet the drug court's eligibility criteria but were arrested in a nearby and demographically similar jurisdiction that does not have a drug court. For example, defendants in two neighboring rural counties within the same state, one with a drug court and one without, may be compared in this fashion. The principal

disadvantage comes from the possibility that local police and prosecutorial practices in the two jurisdictions may differ. This may have a huge impact on the probability that someone is arrested (and charged) for a particular crime. This is especially the case with drug-related crimes, which generally depend upon active police deployment and enforcement activity. For this reason, a comparison jurisdiction approach is not generally preferred. The drug court staff plays a critical role in helping the evaluator to determine whether or not a truly comparable jurisdiction in fact exists.

Ineligible for Drug Court. An ineligible comparison group would consist of defendants considered for drug court participation but found ineligible. The assigned prosecutor may have decided the alleged crimes were too serious to merit the drug court opportunity, the case may have been referred to standard probation instead of drug court, or the defendant may not have been assessed as drug-addicted. The reasons for ineligibility would probably lead ineligible defendants to differ from real participants in important ways. A sole exception might be in drug courts where staff has good reason to believe that many defendants are being found ineligible for wholly arbitrary reasons. In general, however, this approach has significant shortcomings.

Drug Court Failures. A small number of completed studies attempt to demonstrate drug court success by comparing the recidivism rates of graduates and failures. As discussed above, this approach generates little more than a statement of the obvious: those who enter a program and do well (graduates) have better outcomes than those who enter and do poorly (failures). The responsibility of an evaluation is to show whether a program was successful in general with all of those it intended to treat in the first place.

Conclusion: *An appropriate comparison group consists of defendants who did not participate in the drug court but are similar in other ways. There are a large*

number of potential comparison group designs, each with specific advantages and disadvantages; drug court staff can play a critical role in helping to determine the best and most practical approach. To help the evaluator, staff should carefully review how defendants are routed to the drug court and whether a similar pool exists that is technically eligible but not *routed to the drug court for logistic, bureaucratic, or other unintentional reasons. If there is no such pool arrested during the* same *period of time, staff might recommend drawing the comparison group from defendants arrested before the drug court opened (the "pre-post" approach). Staff should also feel empowered to impart advice on other options (e.g., by explaining the most common reasons for why certain defendants may refuse treatment or by commenting on the potential comparability of nearby jurisdictions that do not have a drug court).*

How Do You Ensure that the Drug Court and Comparison Samples Are Truly Comparable?

[18] Having identified the drug court and comparison samples, the next step is to compare them on all available background characteristics to verify that they are indeed comparable. Ideally, data will be collected on enough key characteristics to avoid the possibility that important "unobserved" differences may still exist. For example, as discussed above, refused treatment comparison groups are often a poor choice due to the possibility that they may differ from drug court participants on characteristics that are usually "unobserved" or unavailable in the data, such as defendant motivation to change their lifestyle.

A trained evaluator will conduct statistical tests to see if "statistically significant" differences exist between the background characteristics of the drug court and comparison samples (e.g., the number of priors, arrest charges, age, race, employment status, drug of choice, and treatment history, to the extent that this data is available). Not all differences need

be a cause of concern. For example, if the average age is 30 for drug court participants and 31 for the comparison group, these numbers are different, but the difference is probably insignificant statistically. Also, it may not be a major problem if the two samples are compared on a large number of characteristics and there are only one or two differences. In general, differences on kinds of characteristics that are likely to affect the probability of recidivism are the most troubling kind. As examples, since younger defendants and defendants with more priors are almost always more likely to re-offend, it is extremely desirable to end up with comparable samples on age and criminal history.

What if the samples are different? All is not lost, because a variety of statistical methods can "control for" or take into account those differences. While it is beyond the purview of this paper to describe the underlying statistics, a few examples are briefly outlined. In general, in working with a trained evaluator, staff should at least feel comfortable asking if the drug court and comparison samples turned out to be comparable and, if they did not, what the evaluator did to correct for any potential biases. The evaluator should be able to produce a few simple charts or descriptions that convey a basic sense of how the evaluator proceeded.

Statistical Controls. Methods known by such terms as "multivariate" or "regression" can be used to determine whether an intervention (i.e., drug court) affects an outcome (i.e., recidivism), after controlling simultaneously, within a single mathematical computation, for the effects of other characteristics (e.g., criminal history, age, race, sex, and so forth). Unfortunately, from the perspective of drug court staff, what is often disappointing about these methods is that they fail to yield simple percentages that are meaningful to the lay reader; while these methods can clearly indicate whether or not the drug court produced a statistically significant reduction in recidivism, to quantify the exact *extent* of the reduction, the method yields raw numbers that,

while they make sense to researchers, lack the transparency of a simple comparison of re-arrest rates (e.g., 50 percent versus 40 percent, 50 percent versus 30 percent, etc.).

Predicted Probabilities. A predicted probability is a probability or percent (e.g., 10 percent, 20 percent, 30 percent, etc.), which is computed *after and in light of* statistical controls. Essentially, the idea is to use the results of the analyses falling under method #1 to determine what the drug court and comparison group recidivism rates *would probably be* if all other characteristics were set to their averages. For example, if the drug court and comparison samples have a combined average age of 30, a 40 percent average probability of being female, a 60 percent probability of having a prior conviction, and so forth, then we can compute, for a hypothetical defendant possessing all of the various average characteristics, what would be the probability of recidivism if that defendant was in the drug court as opposed to the comparison group. While this method can yield simple percentages that are readily comprehensible to the lay reader, the results have a somewhat artificial or "made up" quality, in that few real defendants are entirely "average"; further, the drug court may produce a relatively greater or lesser impact on recidivism for defendants at the extremes (e.g., for extremely young or extremely old defendants) than for those at the average; but this possibility is occluded by the predicted probability approach.

Propensity Scores. This refers to an increasingly popular method, for which there are a large number of permutations. In the most understandable of these methods, evaluators compare the complete set of background characteristics of both the drug court and comparison samples and remove from the final comparison sample defendants whose characteristics comprise a "poor match" to those in the drug court. How is this done? First, a mathematical computation is performed that leads each defendant to be assigned a "propensity score," which essentially represents

the probability that, given the defendant's particular panoply of background characteristics, the defendant would have entered the drug court if the opportunity was available. For instance, potential comparison group defendants in a "pre-post" design obviously did not enter the drug court for the simple reason that it had not opened when the defendants were arrested, but the propensity score serves as a probability that, if the drug court *had* been open, the defendant would have participated. Conversely, all drug court participants obviously did enroll, but some may still have a higher propensity score than others—in other words, they may possess characteristics that led enrollment to have been probabilistically more likely from the outset. Having assigned propensity scores to all defendants, evaluators match, one at a time, each drug court participant to the specific comparison group defendant with the nearest score—i.e., with the most comparable set of background characteristics. Then the evaluators delete from the final sample all comparison defendants for whom a match was not found. The process removes from the final comparison sample all of the poor matches. If the process works, it leaves the analyst with two samples whose background characteristics no longer differ. From there, the analysis can proceed in a straightforward manner, as simple recidivism percentages can be computed and compared between the samples.

Subgroup Analysis. This method can be useful if the initial drug court and comparison samples differ enormously on just one or two key characteristics. To offer a hypothetical, let us suppose that 70 percent of the drug court participant sample is female but only 40 percent of comparison group sample is. One could address this problem by dividing the samples into women and men, and then comparing drug court and comparison group recidivism rates separately for each sex. Perhaps the drug court produces a substantial recidivism reduction for women (e.g., 50 percent to 25 percent) but a smaller one for men (e.g., 60 percent to

Case-in-Point: The Los Angeles County Drug Court Evaluation: In evaluating the Los Angeles County Drug Courts, the researchers identified two types of comparison groups, one consisting of defendants who enrolled in an alternative 20-week education and rehabilitation diversion program (i.e., not the drug court), and a second comparison group consisting of defendants not enrolling in any court-mandated treatment program (Fielding et al., 2002). The researchers then encountered the problem that the average risk level of the three samples (drug court, alternative diversion program, and no-treatment) varied widely—with risk defined by the defendant's prior criminal record, seriousness of the current charges, and community ties. For instance, 29 percent of the drug court sample, a mere 10 percent of the first comparison sample, and a far higher 72 percent of the second comparison sample was classified as "high" or "very high" risk. Since risk level may predict recidivism (e.g., one might expect high-risk defendants to be more likely to re-offend in general), these differences represented an extremely serious source of bias. The researchers solved this problem by reporting all of their key recidivism results separately for subgroups classified into three risk levels: (1) low, (2) medium and (3) high/very high. Using this strategy, they produced the interesting finding that the Los Angeles Drug Courts worked best with medium and high-risk defendants. There were no significant differences in re-arrest rates over a one-year post-program period for those in the low risk category, but the re-arrest rates for those in the medium and high/very high risk categories were significantly lower among drug court participants than among defendants in either of the two comparison groups. Considering the "high/very high" risk category, for example, the re-arrest rate was 21 percent for participants in the drug court, 37 percent for participants in the 20-week alternative diversion program, and 55 percent for defendants not mandated to any treatment-based intervention.

50 percent), which would itself be an interesting finding. One cautionary note with respect to this type of analysis has to do with sample size. By splitting the samples (e.g., into women and men), it will become more difficult to show statistically significant effects (recalling the earlier power analysis discussion). For example, what were once samples of 200 participants and 200 comparison defendants may become samples of only 100 for each sex, which may no longer be sufficient to produce differences in recidivism rates that fall outside the study's margin of error.

Conclusion: *Having identified what appear to be appropriate drug court and comparison group samples, it is still necessary to verify that their background characteristics are indeed similar (e.g., by comparing their demographics, charges, criminal history, and other characteristics). An informed staff can ask the evaluator whether appropriate checks were conducted and can ask for a lay description of what, if any, methods were used to correct for any differences that may have been detected. This communication process between staff and evaluator will increase the confidence of both parties in the ensuing results.*

SUMMARY

Drug court administrators and staff are likely to have superior knowledge of both their own program and of important criminal justice policies in their jurisdiction. For this reason, this paper argues that an informed staff can provide valuable contributions to evaluators trying to sort through key methodological challenges—most importantly of all, the choice of an appropriate and readily available comparison group. For instance, is a strong contemporaneous comparison group possible in the jurisdiction, or are all technically eligible defendants routed straight to the drug court, leaving no one left to include in the comparison group? Can a "pre-post" design be implemented by including in the comparison group defendants who are technically eligible but

who were arrested prior to the opening of the drug court? With greater understanding of basic methodological tools, drug court administrators and staff are fully capable of helping researchers to weigh options and make informed decisions. With a stronger partnership between drug court practitioners and researchers, the quality of the resulting evaluations is sure to improve.

REFERENCES

Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (2001). *The comparative costs and benefits of programs to reduce crime version 4.0*. Olympia, WA: Washington State Institute for Public Policy.

Belenko, S. (2001). *Research on drug courts: A critical review: 2001 update*. New York: National Center on Addiction and Substance Abuse at Columbia University.

Fielding, J.E., Tye, G., Ogawa, P., Imam, I.J., & Long, A.M. (2002). Los Angeles County drug court programs: Initial results. *Journal of Substance Abuse Treatment 23*, 217-224.

Goldkamp, J.S., White, M.D., & Robinson, J. B. (2001). *From whether to how drug courts work: Retrospective evaluation of drug courts in Clark County (Las Vegas) and Multnomah County (Portland)*. Philadelphia, PA: Crime and Justice Research Institute.

Gottfredson, D.C., Najaka, S.S. & Kearley, B. (2003). Effectiveness of drug treatment courts: Evidence from a randomized trial. *Criminology and Public Policy 2*, 171-196.

Gottfredson, D.C., Kearley, B., Najaka, S.S., & Rocha, C. (2003, January). *Baltimore City Drug Treatment Court:Evaluation of client self-reports at three-year follow-up*. Draft M.S.

Marlowe, D.B., D.S. DeMatteo, & D.S. Festinger. (2003). A sober assessment of drug courts. *Federal Sentencing Reporter 16*(1), 113-128.

Marlowe, D.B., Festinger, D.S., & Lee, P.A. (2004). The judge is a key component of drug court. *Drug Court Review IV*(2), 1-34.

Rempel, M. & DeStefano, C.D. (2001). Predictors of engagement in court-mandated treatment: Findings at the Brooklyn Treatment Court, 1996-2000. *Journal of Offender Rehabilitation 33*, 87-124.

Rempel, M., Fox-Kralstein, D., Cissner, A., Cohen, R., Labriola, M., Farole, D., Bader, A., & Magnani, M. (2003). *The New York State adult drug court evaluation: Policies, participants, and impacts*. Report submitted to the New York State Unified Court System and the Bureau of Justice Assistance. New York: Center for Court Innovation.

Roman, J. & DeStefano, C.D. (2004). Drug court effects and the quality of existing evidence. In J. Butts and J. Roman (Eds.), *Juvenile Drug Courts and Teen Substance Abuse*. Washington, DC: Urban Institute Press.

Truitt, L., Rhodes, W.M., Seeherman, A.M., Carrigan, K., & Finn, P. (2002). *Phase I: Case studies and impact evaluations of Escambia County, Florida and Jackson County, Missouri Drug Courts*. Cambridge, MA: Abt Associates.

Wilson, D.B., Mitchell, O., & MacKenzie, D.L. (2002). *A systematic review of drug court effects on recidivism*. Draft M.S.

U.S. Government Accountability Office. (2005, February). *Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes*. Report to congressional committees. Washington, DC: Author.

SUBJECT INDEX

The following cumulative Subject Index is designed to provide easy access to subject references. Each reference can be located by:

- *Volume* *by using a roman numeral e.g. I*
- *Issue* *by using a number e.g. 2*
- *Subject reference* *by its page number in parenthesis e.g. (121)*

HEADNOTE INDEX

The Headnote Index provides access to an article's major points or concepts using a cumulative indexing system. Each headnote can be located by:

- *Volume* *by using a roman numeral i.e. I*
- *Issue* *by using a number i.e. 2*
- *Headnote* *by using a number in brackets i.e. [4]*

Perceptions of Drug Court

II1**[15]** Evaluating the FTDO Program in Maricopa
II1**[16]** 12-Month/36-Month Outcomes
II1**[17]** Difficulty of Compliance
II1**[18]** Helpfulness, Strengths/Weakness

Performance Measurement

V2**[5]** What Is Performance Measurement?
V2**[6]** Measuring Drug Court Performance
V2**[7]** Conclusion

Process Evaluation

V2**[8]** What Are Process Evaluations?
V2**[9]** Who Should Conduct These Evaluations?
V2**[10]** What Are the Critical Elements of Process Evaluation?
V2**[11]** What Data Are Needed to Complete These Evaluations?
V2**[12]** What Does It Mean to be "Methodologically Rigorous"?
V2**[13]** What About Experimental Design and Comparison Groups?

Recidivism

V2**[14]** What Do We Know Now?
V2**[15]** What Is "Recidivism"?
V2**[16]** Which Drug Court Participants Should Be Included in the Analysis?
V2**[17]** What Is an Appropriate Comparison Group?
V2**[18]** How Do You Ensure that the Drug Court and Comparison Samples Are Truly Comparable?

Research

II1**[27]** Recidivism and The Utah Juvenile Court
II1**[28]** Delaware Drug Court Evaluation
II1**[29]** Florida's First Judicial Circuit Drug Court Evaluation
II1**[30]** Monterey County First-Year Drug Court Evaluation
II1**[31]** Riverside County Drug Court Evaluation
II2**[21]** Monterey County, 1st Year Evaluation
II2**[22]** Butler County CDAT Evaluation
II2**[23]** King County Drug Court Evaluation
II2**[24]** Suffolk County Drug Treatment Court
II2**[25]** Volusia County Process & Output Evaluation
II2**[26]** Jefferson County Impact Evaluation
II2**[27]** Madison County Final Evaluation
II2**[28]** Santa Barbara County Year Three
III1**[25]** Cleveland Drug Court
III1**[26]** Allen County
III1**[27]** Delaware Juvenile Diversion Program
III1**[28]** Orange County
III1**[29]** Creek County
III1**[30]** Project Exodus (Maine)
III1**[31]** Denver Drug Court
III2**[32]** Dallas County DIVERT Court
III2**[33]** Maine's State-wide Adult Drug Treatment Court Program
III2**[34]** Maine's State-wide Juvenile Drug Treatment Court Program