

## **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 8<sup>th</sup> Avenue, 18<sup>th</sup> Floor, New York, New York 10018. (212) 373-1681; (212) 397-0985 (fax); [rempelm@courtinnovation.org](mailto: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

**A**midst 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.