CENTER FOR COURT INNOVATION
A Public/Private Partnership with the
New York State Unified Court System

RESEARCH

Recidivism 101

Evaluating the Impact of Your Drug Court

Written by

Michael Rempel

2005

This publication was supported by the Bureau of Justice Assistance, Office of Justice Programs, U.S. Department of Justice under Grant Number 98-DC-VX-K007 awarded to the Center for Court Innovation. Any opinions and interpretations expressed in this paper are those of the author and do not necessarily represent the official position of the U.S. Department of Justice.

About the Author

Michael Rempel is research director at the Center for Court Innovation.

Acknowledgements

The author would like to thank Donald J. Farole, Jr., Greg Berman, Valerie Raine, and Robert V. Wolf at the Center for Court Innovation for their helpful feedback and guidance in writing this paper.

Evaluating the Impact of Your Drug Court

Introduction

Amidst widespread agreement that producing reductions in recidivism is a universal drug court goal, court administrators and drug court 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.

This paper provides a lay summary of the four core methodological questions that must be addressed in any recidivism analysis. Real examples from the evaluation literature are incorporated throughout to show how different methods have been applied. For overview purposes, the four questions are:

1. Which drug court participants should you include 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?

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

3. 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?

4. What is the definition of recidivism?
What recidivism measures are appropriate (e.g., re-arrest, reconviction, re-incarceration, or other measures)? What is the ideal timeframe to measure recidivism (e.g., one year after drug court participation begins, two years after participation begins, one year after program exit, etc.)?

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?

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. For instance, in a comprehensive 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 studied; 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.[1]

While this review is extremely positive, much of the recidivism literature, particularly the first generation of studies completed in the 1990s, possessed serious methodological shortcomings.[2] Most notable was a failure to identify an appropriate "comparison group" of defendants with whom drug court recidivism rates could be reasonably compared. 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 very recent literature 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.[3]

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 inclusion in the comparison group. While only a handful isolated post-program recidivism (after participants have either graduated or failed), results from these studies too are encouraging. For example, 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.[4] A rigorous study of the Los Angeles County Drug Courts similarly isolated recidivism during a one-year post-program period. Interestingly, this study found that the drug court produced significant reductions in recidivism 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).[5] Several other studies have confirmed that various aspects of the drug court model work particularly well with high-risk defendants.[6]

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, Doug Marlowe 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."[7] 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 perfectly 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.

Question One: Which Drug Court Participants Should You Include in the Analysis?

The first step in any 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. A sample size of at least 100-200 participants, if not more, is probably necessary to generate "statistically significant" results that fall outside the study's margin of error. The general rule is that the greater the sample size, the smaller the study's margin of error, although adding more sample size is far more helpful at the low end of the spectrum (e.g., going from 100 to 200 participants) than at the high end (e.g., going from 400 to 500 participants).

In most adult drug courts, identifying the participants is fairly 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 still 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 any evaluative significance. As a policy matter, what is important to know 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 who had 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, or working out kinks in its operations. For instance, in evaluating the Rochester (New York) Drug Treatment Court, because it was the first to open in New York State and had to develop much of its model after operations began, the researchers decided to exclude drug court participants enrolling in 1995, the first year of operations.[8] Second, one always excludes the most recent entrants, since they will not have been in the program for long enough to enable analyzing their recidivism rates. 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). Third, when attempting to analyze recidivism over a post-program period after graduation or failure, it goes without saying that only graduates and failures should be included, not participants with unresolved cases. Finally, some participants may need to be excluded due to missing data on key variables. The general goal is to obtain a "representative" sample, but not necessarily a complete one. Of course, if key categories of participants systematically have missing data more often than others (e.g., if failures have missing data more often than graduates), that problem would be more serious and could lead to biased results. (Such a problem is still correctable through weighting techniques, whereby the analysis would value the recidivism results of participants in underrepresented categories more than participants in the overrepresented categories.)

A separate consideration is the number of available participants. There is no point in examining whether a drug court reduced the recidivism of its first 30 participants, since that sample size is too small to yield meaningful results. What sample size is enough? A "power analysis" is a common method researchers use to answer this question. 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. A difference of this magnitude would be close to unprecedented in the literature. Although some drug courts have been able to achieve the impact that would be required with samples 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 possible for the average successful program actually to show a 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 samples from 600 to 800.

Comparison Group Sample Size	Drug Court Sample Size	Comparison Group Re-Arrest Rate	Drug Court Re-Arrest Rate Needed to Detect Significant Effect
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

Conclusion: A representative sample of drug court participants (including both graduates and failures) should be included. Since a recidivism analysis is unlikely to produce meaningful results with fewer than 100-200 participants, drug court staff should communicate its rate of intake to the evaluator and help the evaluator develop a realistic timeline for conducting the analysis.

Question Two: What Is an Appropriate Comparison Group?

The performance of drug court participants becomes meaningful only in relation to a "comparison group": defendants who did not enter drug court but are similar in their criminal justice status and other characteristics (e.g., demographics, substance abuse history). It is important for the background characteristics of the comparison group to be as similar as possible to those of the drug court participants; otherwise, the recidivism results may be misleading. To illustrate why, consider the implications of having dissimilar samples with respect to prior criminal history. It is well known that defendants with more prior offenses are generally more likely to commit future offenses. Therefore, if the drug court sample averages fewer priors than the comparison group, and if drug court 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).

Random Assignment

This is the gold standard. First, defendants are screened to determine whether they are eligible for the drug court and willing to participate. Then those who are eligible and willing are randomly assigned to either the drug court or the comparison group. In theory, random assignment 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 the random assignment. In practice, these designs are extremely rare for ethical reasons—it requires denying a treatment thought to be effective to the comparison group.

Also, random assignment studies often face serious implementation problems. For example, the research integrity of random assignment studies may be compromised if judges or other court staff can selectively remove large numbers of defendants from the random assignment process; or if the drug court changes its eligibility criteria as a result of the study, for example by allowing defendants arrested only on less serious charges to participate in the random assignment. Therefore, while in the abstract random assignment is the best method available, it often encounters important limitations. Moreover, these designs are not the only ones able to yield valid results.

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 same period of time. 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 not found to be drug-addicted? Did they refuse to participate? Or did

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 the assignment took place, 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 an extremely 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. [9]

Case-in-Point: The Brooklyn and Rochester Drug Court Evaluations

The statewide evaluation of New York's drug courts included recidivism studies of six separate sites, and in two of these, it was possible to identify a strong contemporaneous comparison group. For the evaluation of the Brooklyn Treatment Court, the researchers took advantage of a situation in which the drug court only accepted defendants arrested in three of five geographic arrest zones of Brooklyn; defendants arrested in the two remaining zones and who otherwise met the drug court's eligibility criteria comprised the comparison group. For the 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; the comparison group thus consisted of defendants arraigned on drug court-eligible charges by a judge other than those two. Both evaluations showed small but significant drug court impacts over a one-year post-program period of time. In Brooklyn, the reconviction rate for drug court participants was 17 percent compared with 23 percent for the comparison group; and in Rochester, the reconviction rate was 42 percent compared with 48 percent for the comparison group. [10]

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 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 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 (based on their formal charges and criminal history), 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.

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 the potential comparison group defendants are drug-addicted. Instead, the comparability of participants to comparison 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 in the period either before or 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 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.

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 has a critical shortcoming, since those unwilling to enter the drug court clearly differ from real participants: i.e., by definition, they lack interest or motivation to participate. 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 this comparison group may certainly be used when it is the only one feasible, evaluators should probably avoid it if a better design is possible and should take extra care to investigate if or how it may be biased. This may be done through statistical methods briefly outlined below or through qualitative interviews with case managers or drug court staff designed to probe why some defendants opt not to participate. (And drug court staff can assist the evaluator here by volunteering information about why some defendants refuse.)

In general, while some may be concerned about the problem of historical bias when using the previously described "pre-post" design, it is this author's opinion that a "pre-post" design is usually preferable to a "refused treatment" design whenever both are feasible options. The act of refusing treatment makes every single defendant in such a comparison group necessarily different from real drug court participants,

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.[11]

whereas defendants with what appear to be similar characteristics who were arrested prior to the opening of a drug court are quite likely, indeed, to make for a perfectly valid comparison. Realistically, police and 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 if available.

Comparison Jurisdiction

This type of comparison group consists of defendants who meet the drug court's eligibility criteria but were arrested in a nearby 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 reality that police and prosecutorial practices may differ from one jurisdiction to the next. 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 are often unreported and therefore depend upon active police enforcement. For this reason, while a comparison jurisdiction is not generally preferred, it can work under the right circumstances; and court staff can play 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 the drug court; or the defendant may not have been found drug-addicted. The reasons for ineligibility would probably lead ineligible defendants to differ from real participants. 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, much like a refused treatment comparison group, this one has significant shortcomings.

Aggregate Data

Typically, states and counties compute aggregate recidivism rates for various categories of offenders (e.g., drug offenders, property offenders, violent offenders, those placed on probation, those placed on parole). One might use these aggregate recidivism rates as a baseline, against which to compare the performance of drug court participants. Doing so can be perfectly useful and worthwhile as an easy way for drug court staff to gain a rough sense of how participants are faring. But this method cannot yield rigorous results and should never be used in a formal evaluation. It is impossible to know how similar or different from real drug court participants are those encompassed in aggregate data. Indeed, since all drug courts have specific screening and assessment policies leading participants with certain characteristics to be particularly likely to be found ineligible (for legal or clinical reasons), the characteristics (e.g., criminal history, age, other demographics, substance abuse history) of the average real participant probably differs greatly from those of the wider pool of defendants included in various aggregate statistics. To ensure that a recidivism analysis is comparing apples to apples, a specific comparison group is necessary: i.e., a set of specific defendants, for which at least the most basic criminal justice and demographic characteristics are known.

Drug Court Failures

A small number of completed studies attempt to demonstrate drug court success by comparing the recidivism rates of graduates to 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 all other ways. There are a large number of potential comparison group designs, each with specific advantages and disadvantages; and drug court staff can play a critical role in helping to determine the best available 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).

Question Three: How Do You Ensure That the Final Drug Court and Comparison Samples Are Truly Comparable?

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 usually 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 future behavior.

A trained evaluator will conduct statistical tests, such as a "T test," to see if statistically significant differences exist between the drug court and comparison samples (e.g., in the average number of priors, arrest charges, race, average age, employment status, and drug of choice, 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, if the two samples are compared on a large number of characteristics and there are one or two differences, that may not be a major problem. In general, differences on kinds of characteristics that are likely to affect the probability of recidivism are the most troubling kind. In particular, 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 in fact 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

"Multivariate" or "regression" methods can be used to determine whether an intervention (e.g., drug court) affects an outcome (e.g., re-arrest), after controlling simultaneously, within a single mathematical computation, for the effects of other characteristics (e.g., criminal history, age, race, sex, etc.). 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 drug court participation produced a statistically significant reduction in recidivism, to quantify the exact extent of the reduction, the method yields numbers that, while they make sense to researchers, lack the transparency of a simple comparison of re-arrest rates. For example, "odds ratios" of 1.5 or 2.3 make less intuitive sense to policymakers than a recidivism reduction of 50 percent versus 40 percent or 50 percent versus 30 percent.

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 analyses falling under the statistical controls method 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 average 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 Score Matching

This is an increasingly popular method for which there are a large number of permutations. In the simplest of these permutations, the method involves comparing the complete set of background characteristics of both the drug court and comparison samples and removing 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 yields 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 made enrollment 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; they then delete all comparison defendants for whom a match was not found. The process removes from the final comparison sample all of the poor matches. Therefore, if the process works, it leaves the analyst with two samples whose background characteristics do not differ anymore. From there, the analysis can proceed in a straightforward manner, as simple recidivism percentages can be computed and compared between the samples.[12] A propensity score matching strategy was used in the six-site New York State evaluation. In five sites, multiple significant differences distinguished the characteristics of the initial drug court and comparison samples; but these differences were vastly reduced after the matching process.

Subgroup Analysis

This method can be useful if the initial drug court and comparison samples differ widely 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 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.

Case-in-Point:
The Los Angeles County Drug Court Evaluation

In evaluating the Los Angeles 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. 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.[13]

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.

Question Four: What Is the Definition of Recidivism?

Most completed drug court evaluations define recidivism based on re-arrests; some also use reconvictions instead or in addition.[14] Re-arrest-based measures are often preferred for a couple of 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 and re-arrests take within a much shorter timeframe. 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 have subsequently been established.

In drug courts that accept only 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 in the 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.[15]

More important than the choice of measure (re-arrest or reconviction) 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.[16]

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 for a post-program recidivism analysis to begin. Therefore, if drug court staff would like to see some recidivism results on a more timely schedule, that would argue for foregoing, or at least postponing, a post-program analysis in favor of an in-program one.

Whatever timeframe is selected, drug court participants and the comparison group should be analyzed similarly. 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 date of drug court failure or of final disposition on the initial case, but on the date of release from incarceration in the event that the defendant was sentenced to jail or prison; this is to avoid the obvious finding that defendants were not re-arrested while they were incarcerated.

Conclusion: Recidivism is usually defined as "re-arrest" but sometimes as reconviction; 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). Drug court staff should feel free to discuss with the evaluator its own preferences for defining recidivism. Furthermore, staff should express its preferences for the analysis timeframe, recognizing the tradeoff that 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.

Summary

For the lay practitioner, it is natural for issues of recidivism methodology to fall somewhere between daunting and downright impenetrable. One solution is to avoid such issues altogether, leaving them for trained evaluators, who are paid as an express result of their purported skill at knowing how to conduct methodologically sound recidivism studies. And yet, as a series of recent drug court literature reviews have found, a great many completed drug court evaluations, especially among the first generation of studies from the 1990s, were hindered by serious methodological shortcomings. So perhaps trained researchers *cannot* always be relied upon to figure out how to conduct their evaluation while steering clear of methodological pitfalls.

Since drug court administrators and staff are likely to have superior knowledge of both their own drug court program and of criminal justice policies in their jurisdiction, this paper argues that an informed staff can make valuable contributions to evaluators trying to sort through important 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 would have been eligible but who were arrested prior to the opening of the drug court? With a greater understanding of basic methodological tools, drug court administrators are fully capable of helping researchers weigh options and make informed decisions. With a stronger partnership between drug court staff and researchers, the quality of the resulting recidivism studies is sure to improve.

Notes

1. See D. B. Wilson, O. Mitchell, and D. L. MacKenzie, "A Systematic Review of Drug Court Effects on Recidivism," Draft M.S. (2002).
2. See critiques in S. Belenko, "Research on Drug Courts: A Critical Review: 2001 Update." New York: National Center on Addiction and Substance Abuse at Columbia University (2001); and J. Roman and C. DeStefano, "Drug Court Effects and the Quality of Existing Evidence," In, *Juvenile Drug Courts and Teen Substance Abuse*, eds. J. Butts and J. Roman. Washington, DC: Urban Institute Press (2004).

3. See S. Aos, P. Phipps, R. Barnoski, and R. Lieb, *The Comparative Costs and Benefits of Programs to Reduce Crime Version 4.0.* Olympia: Washington State Institute for Public Policy (2001); Government Accountability Office, *Adult Drug Courts: Evidence Indicates Recidivism Reductions and Mixed Results for Other Outcomes,* United States Government Accountability Office, Report to Congressional Committees (February 2005); and J. Roman and C. DeStefano, *supra* note 2.
4. M. Rempel, D. Fox-Kralstein, A. Cissner, R. Cohen, M. Labriola, D. Farole, A. Bader, and M. Magnani. *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 (2003).
5. J. E. Fielding, G. Tye, P. Ogawa, I. J. Imam, A. M. Long, "Los Angeles County Drug Court Programs: Initial Results," 23 *Journal of Substance Abuse Treatment,* 217, 224 (2002).
6. See, for example, D. B. Marlowe, D. S. Festinger, and P. A. Lee, "The Judge is a Key Component of Drug Court," *National Drug Court Institute Review* IV: 2: 1-34 (2004); and M. Rempel and C. D. DeStefano, "Predictors of Engagement in Court-Mandated Treatment: Findings at the Brooklyn Treatment Court, 1996-2000" 33 *Journal of Offender Rehabilitation,* 87, 124 (2001).
7. D. Marlowe, "A Sober Assessment of Drug Courts," *Federal Sentencing Reporter* 16:1 (October 2003).
8. M. Rempel et al., *supra* note 4, see Chapter 17.
9. See D. C. Gottfredson, S. S. Najaka, and B. Kearley, "Effectiveness of Drug Treatment Courts: Evidence from a Randomized Trial," *Criminology and Public Policy* 2: 171-196 (2003); and D. C. Gottfredon, S. S. Najaka, B. W. Kearley, and C. Rocha, "A Randomized Study of the Baltimore City Drug Treatment Court: Results from the Three-Year Follow-Up." Paper presented at the Annual Conference of the American Society of Criminology, Chicago, IL (2002).
10. M. Rempel et al, *supra* note 4, see chapters 13 (Brooklyn) and 17 (Rochester).
11. M. Rempel et al, *supra* note 4, see chapter 12.
12. A more detailed discussion of this technique as applied in a drug court evaluation can be found in Rempel et al., *supra* note 4, chapter 11; with applications to six specific courts discussed in chapters 12-17.
13. J. E. Fielding, et al., *supra* note 5.
14. See studies reviewed in D. B. Wilson et al., *supra* note 1 or Gov. Accountability Office, *supra* note 2.
15. L. Truitt, W. M. Rhodes, A. M. Seeherman, K. Carrigan, and P. Finn. *Phase I: Case Studies and Impact Evaluations of Escambia County, Florida and Jackson County, Missouri Drug Courts,* Abt Associates (March 2000).
16. See discussion in S. Belenko, *supra* note 2.

Center for Court Innovation

The winner of an Innovations in American Government Award from the Ford Foundation and Harvard's John F. Kennedy School of Government, the Center for Court Innovation is a unique public-private partnership that promotes new thinking about how courts and criminal justice agencies can aid victims, change the behavior of offenders and strengthen communities.

In New York, the Center functions as the state court system's independent research and development arm, creating demonstration projects that test new approaches to problems that have resisted conventional solutions. The Center’s problem-solving courts include the Red Hook Community Justice Center and the Midtown Community Court as well as drug courts, domestic violence courts, youth courts and mental health courts.

Nationally, the Center disseminates the lessons learned from its experiments in New York, helping practitioners across the country launch their own problem-solving innovations. The Center contributes to the national conversation about justice through a variety of written products, including original research, books and white papers like this one. The Center also provides hands-on technical assistance, advising courts, prosecutors and other criminal justice planners throughout the country.

For more information, call 212 397 3050 or e-mail info@courtinnovation.org.

Center for Court Innovation
520 Eighth Avenue, 18th Floor
New York, New York 10018
212 397 3050 Fax 212 397 0985
www.courtinnovation.org

CENTER
FOR
COURT
INNOVATION

A Public/Private Partnership with the
New York State Unified Court System