X. MONITORING AND EVALUATION

The Drug Court routinely monitors its adherence to best practice standards and employs scientifically valid and reliable procedures to evaluate its effectiveness.[8]

A. Adherence to Best Practices
B. In-Program Outcomes
C. Criminal Recidivism
D. Historically Discriminated Against Groups
F. Electronic Database
G. Timely and Reliable Data Entry
H. Intent-to-Treat Analyses
I. Comparison Groups
J. Time at Risk

A. Adherence to Best Practices

The Drug Court monitors its adherence to best practice standards on at least an annual basis, develops a remedial action plan and timetable to rectify deficiencies, and examines the success of the remedial actions. Outcome evaluations describe the effectiveness of the Drug Court in the context of its adherence to best practices.

B. In-Program Outcomes

The Drug Court continually monitors participant outcomes during enrollment in the program, including attendance at scheduled appointments, drug and alcohol test results, graduation rates, lengths of stay, and in-program technical violations[9] and new arrests.

C. Criminal Recidivism

Where such information is available, new arrests, new convictions, and new incarcerations are monitored for at least three years following each participant's entry into the Drug Court. Offenses are categorized according to the level (felony, misdemeanor, or summary offense) and nature (e.g., person, property, drug, or traffic offense) of the crime involved.

---

[8] Herein, monitoring refers to periodic descriptions of the services delivered and outcomes achieved in a Drug Court without inferring a causal relationship between the services and outcomes. An evaluation includes a comparison condition and other scientific procedures designed to attribute outcomes to the effects of the Drug Court. Most Drug Courts are capable of monitoring their services and outcomes but may require expert consultation to evaluate the causal effects of their program.

[9] A *technical violation* refers to a violation of a court order that does not constitute a crime per se. For example, drinking alcohol is legal for most adults but is usually a technical violation in a Drug Court.

D. Independent Evaluations

A skilled and independent evaluator examines the Drug Court's adherence to best practices and participant outcomes no less frequently than every five years. The Drug Court develops a remedial action plan and timetable to implement recommendations from the evaluator to improve the program's adherence to best practices.

E. Historically Discriminated Against Groups

The Drug Court continually monitors admission rates, services delivered, and outcomes achieved for members of groups that have historically experienced discrimination who are represented in the Drug Court population. The Drug Court develops a remedial action plan and timetable to correct disparities and examines the success of the remedial actions [see also Standard II, Equity and Inclusion].

F. Electronic Database

Information relating to the services provided and participants' in-program performance is entered into an electronic database. Statistical summaries from the database provide staff with real-time information concerning the Drug Court's adherence to best practices and in-program outcomes.

G. Timely and Reliable Data Entry

Staff members are required to record information concerning the provision of services and in-program outcomes within forty-eight hours of the respective events. Timely and reliable data entry is required of each staff member and is a basis for evaluating staff job performance.

H. Intent-to-Treat Analyses

Outcomes are examined for all eligible participants who entered the Drug Court regardless of whether they graduated, withdrew, or were terminated from the program.

I. Comparison Groups

Outcomes for Drug Court participants are compared to those of an unbiased and equivalent comparison group. Individuals in the comparison group satisfy legal and clinical eligibility criteria for participation in the Drug Court, but did not enter the Drug Court for reasons having no relationship to their outcomes. Comparison groups do not include individuals who refused to enter the Drug Court, withdrew or were terminated from the Drug Court, or were denied entry to the Drug Court because of their legal charges, criminal history, or clinical assessment results.

J. Time at Risk

Participants in the Drug Court and comparison groups have an equivalent opportunity to engage in conduct of interest to the evaluation, such as substance use and criminal recidivism. Outcomes for both groups are examined over an equivalent time period beginning from a comparable start date. If participants in either group were incarcerated or detained in a residential facility for a significantly longer period of time than participants

in the other group, the length of time participants were detained or incarcerated is accounted for statistically in outcome comparisons.

COMMENTARY

A. Adherence to Best Practices

Adherence to best practices is generally poor in most sectors of the criminal justice and substance use disorder treatment systems (Friedmann et al., 2007; Henderson et al., 2007; McLellan et al., 2003; Taxman et al., 2007). Programs infrequently deliver services that are proven to be effective and commonly deliver services which have not been subjected to careful scientific scrutiny. Over time, the quality and quantity of the services provided may decline precipitously (Etheridge et al., 1995; Van Wormer, 2010). The best way for a Drug Court to guard against these prevailing destructive pressures is to monitor its operations routinely, compare its performance to established benchmarks, and seek to align itself continually with best practices. Not knowing whether one's Drug Court is in compliance with best practices makes it highly unlikely that needed improvements will be recognized and implemented; therefore, evaluating a Drug Court's adherence to best practice standards is, itself, a best practice.

Studies reveal that Drug Courts are significantly more likely to deliver effective services and produce positive outcomes when they hold themselves accountable for meeting empirically validated benchmarks for success. A multisite study involving approximately seventy Drug Courts found that programs had more than twice the impact on crime and were more than twice as cost-effective when they monitored their operations on a consistent basis, reviewed the findings as a team, and modified their policies and procedures accordingly (Carey et al., 2008, 2012).

Like many complex service organizations, Drug Courts are highly susceptible to *drift*, in which the quality of their services may decline appreciably over time (Van Wormer, 2010). Management strategies such as continuous performance improvement (CPI), continuous quality improvement (CQI), and managing for results (MFR) are designed to avoid drift and enhance a program's adoption of best practices. Each of these management strategies emphasizes continual self-monitoring and rapid-cycle testing. This process involves collecting real-time information about a program's operations and outcomes, feeding that information back to key staff members and decision makers on a routine basis, and implementing and evaluating remedial action plans where indicated. Research consistently shows that continual self-monitoring and rapid-cycle testing are critical elements for improving outcomes and increasing adoption of best practices in the health care and criminal justice systems (Damschroder et al., 2009; Rudes et al., 2013; Taxman & Belenko, 2013). These strategies are essential for programs that require cross collaboration and interdisciplinary communication among multiple service agencies, including Drug Courts (Bryson et al., 2006; Wexler et al., 2012).

Studies have not determined how frequently programs should review performance information and implement and evaluate self-corrective measures. Common practice among successful organizations is to collect performance data continually and meet at least annually as a team to review the information and take self-corrective measures (Carey et al., 2012; Rudes et al., 2013; Taxman & Belenko, 2013).

Reporting outcomes from Drug Courts without placing those findings into context by describing the quality of the programs is no longer enough. Meta-analyses (Aos et al., 2006; Latimer et al., 2006; Lowenkamp et al., 2005; Mitchell et al., 2012; Shaffer, 2010; Wilson et al., 2006) and large-scale multisite studies (Rossman et al., 2011) have already clearly established that Drug Courts reduce crime by approximately 8% to 14% on average. These averages, derived from evaluations of more than 100 Drug Courts, mask a great deal of variability between programs. Some Drug Courts reduce crime by more than 50%, others have no impact on crime, and still others increase crime rates in their communities (Carey et al., 2012; Carey & Waller, 2011; Cissner et al., 2013; Downey & Roman, 2010; Government Accountability Office, 2011; Mitchell et al., 2012; Shaffer, 2010). The important question is no longer whether Drug Courts can work, but rather how

they work and what services contribute to better outcomes (Marlowe et al., 2006). Understanding what distinguishes effective Drug Courts from ineffective and harmful Drug Courts is now an essential goal for the field. Unless evaluators describe each Drug Court's adherence to best practices, there is no way to place that program's outcomes in context or interpret the significance of the findings.

B. In-Program Outcomes

One of the primary aims of a Drug Court is to rehabilitate seriously addicted individuals, which means that retaining participants in treatment, reducing drug and alcohol use, and helping participants to complete treatment successfully are important indicators of short-term progress. However, policymakers, the public, and other stakeholders are likely to judge the merits of a Drug Court by how well it reduces crime, incarceration rates, and taxpayer expenditures. Therefore, Drug Courts need to measure in-program outcomes that not only reflect clinical progress, but are also significant predictors of postprogram criminal recidivism and other long-term outcomes.

At minimal cost and effort, Drug Courts can evaluate short-term outcomes while participants are enrolled in the program. These short-term outcomes provide significant information about participants' clinical progress and the likely long-term impacts of the Drug Court on public health and public safety. Studies have consistently determined that postprogram recidivism is reduced significantly when participants attend more frequent treatment and probation sessions, provide fewer drug-positive urine tests, remain in the program for longer periods of time, have fewer in-program technical violations and arrests for new crimes, and satisfy other conditions for graduation (Gifford et al., 2014; Gottfredson et al., 2007, 2008; Huebner & Cobbina, 2007; Jones & Kemp, 2011; Peters et al., 2002). Drug Courts should, therefore, monitor and report on these in-program outcomes routinely during the course of their operations.

Several resources are available to help Drug Courts define and calculate performance measures of in-program outcomes (Berman et al., 2007; Heck, 2006; Marlowe, in press; Peters, 1996; Rubio et al., 2008a). In 2006, NADCP convened leading Drug Court researchers and evaluators to form the National Research Advisory Committee (NRAC). One goal of this committee was to define a core data set of in-program performance measures for adult Drug Courts (Heck, 2006). NRAC selected measures that are simple and inexpensive to track and evaluate and proven to predict long-term outcomes. These performance measures include the following:

- *Retention*—the number of participants who completed the Drug Court divided by the number who entered the program
- *Sobriety*—the number of negative drug and alcohol tests divided by the total number of tests performed
- *Recidivism*—the number of participants arrested for a new crime divided by the number who entered the program, and the number of participants adjudicated officially for a technical violation divided by the number who entered the program
- *Units of Service*—the numbers of treatment sessions, probation sessions, and court hearings attended
- *Length of Stay*—the number of days from entry to discharge or the participant's last in-person contact with staff

Longer lists of performance measures addressing a wide range of outcomes in Drug Courts and other problem-solving courts have been published by expert organizations including the National Center for State Courts (Rubio et al., 2008a; Waters et al., 2010), the Center for Court Innovation (Rempel, 2006, 2007), American University (Peters, 1996), the Organization of American States (Marlowe, in press), the National Center for DWI Courts (Marlowe, 2010), and the National Institute of Justice (NIJ, 2010). Drug Courts are advised to consult these and other resources for further information on how to calculate and interpret additional performance measures for their evaluations.

C. Criminal Recidivism

For many policymakers and members of the public, reducing criminal recidivism is one of the primary aims of a Drug Court. Recidivism is defined as any return to criminal activity after the participant entered the Drug

Court. Recidivism does not include crimes that occurred before the participant entered Drug Court even if those crimes are charged or prosecuted after entry.

Recidivism is measured most commonly by new arrests, new convictions, or new incarcerations occurring over a two- or three-year period (Carey et al., 2012; King & Elderbroom, 2014; Rempel, 2006). For example, the Bureau of Justice Statistics (BJS) tracks new arrests, convictions, and incarcerations occurring within three years of the date that state and federal inmates are released from jail or prison (Durose et al., 2014).

Based on scientific considerations, evaluators should follow participants for at least three years, and ideally up to five years, from the date of entry into the Drug Court or from the date of the arrest or technical violation that made the individual eligible for Drug Court. The date of entry should be the *latest* start date for the evaluation because that is when the Drug Court becomes capable of influencing participant behavior directly.

Starting from the date of arrest or technical violation takes into account the potential impact of delays in admitting participants to Drug Court. The sooner participants enter Drug Court after an arrest or probation violation, the better the results (Carey et al., 2008, 2012); therefore, evaluators may wish to examine how delayed entry affects outcomes. However, because Drug Courts cannot always control what transpires before participants enter the Drug Court program, attributing to the Drug Court any recidivism occurring before entry may not fairly represent the Drug Courts' effects on recidivism. Starting from the date of entry ensures recidivism may be attributed fairly to the effects of the Drug Court. No one answer fully addresses the issues surrounding selection of a start date for evaluation; therefore; evaluators should state clearly what start date was selected and the rationale for choosing that start date.

Rates of criminal recidivism among drug-involved offenders become relatively stable after approximately three to five years (King & Elderbroom, 2014). After three years, statistically significant between-group differences in recidivism are likely to remain significant going forward (e.g., Knight et al., 1999; Martin et al., 1999; Wexler et al., 1999). For example, if Drug Court participants have significantly lower rearrest rates than comparison group subjects after three years, this difference is likely (although not guaranteed) to remain significant after an additional two years (DeVall et al., 2015). After five years, recidivism rates tend to reach a plateau, meaning that most (but not all) participants who will recidivate have likely done so by then (e.g., Gossop et al., 2005; Inciardi et al., 2004; Olson & Lurigio, 2014).

Importantly, these findings do not suggest Drug Courts must wait three to five years before reporting recidivism outcomes. Recidivism occurring during enrollment and shortly after discharge from Drug Court may be of considerable interest to practitioners, policymakers, and other stakeholders. However, implying that recidivism rates occurring within the first two years are likely to reflect the long-term effects of a Drug Court is inappropriate. Evaluators should state clearly that such recidivism rates are preliminary and likely to increase over time.

No one basis exists for deciding whether new arrests, new convictions, or new incarcerations are likely to be the most valid or informative indicator of recidivism. As discussed below, each measure has advantages and disadvantages that the evaluator must take into account. Because no one measure is clearly superior to another, whenever possible evaluators are advised to report all three measures of recidivism, discuss the implications and limitations of each, or indicate why a particular measure is not being reported.

Analyzing new arrests as a measure of criminal recidivism provides at least two advantages. First, arrests are often substantially closer in time to the alleged offense than convictions. Resolving a criminal case and determining guilt or innocence may take months or years. Evaluators can usually report arrest outcomes in much less time than waiting for lengthy legal proceedings to resolve. Second, criminal cases are often dismissed or pled down to a lesser charge for reasons having little to do with factual guilt, such as insufficient evidence or plea bargains. As a result, the absence of a conviction or conviction on a lesser charge may not reflect the offense that occurred.

However, some individuals are arrested for crimes they did not commit. This fact may lead to an overestimation of the true level of criminal recidivism. Relying on conviction data rather than arrest data may provide greater assurances that the crimes did, in fact, occur.

Incarceration has substantial cost impacts that may far exceed those of arrests and convictions. A day in jail or prison can cost between five and twenty times more than a day on probation or in community-based treatment (Belenko et al., 2005; Zarkin et al., 2012). Evaluators typically distinguish between incarceration that occurred while participants were enrolled in the Drug Court and incarceration that occurred after discharge. In-program incarceration often reflects brief jail sanctions that may be imposed for misconduct in the program, whereas postprogram incarceration typically reflects pretrial detention for new charges, sentences for new charges, or (for terminated participants) sentencing on the original charge that led to participation in Drug Court. In cost evaluations, in-program jail sanctions are typically counted as an investment cost for the Drug Court whereas postprogram detention is typically counted as an outcome cost (Carey et al., 2012).

Evaluators must also consider the timeliness and accuracy of information contained in criminal justice databases. In some jurisdictions, arrest data may be recorded in a more timely and faithful manner than conviction or incarceration data. Evaluators must familiarize themselves with how and when information is entered into national, state, and local criminal justice records and should describe clearly in their evaluation reports any limitations that may relate to the accuracy or timeliness of the data.

Self-report information could potentially provide the most accurate assessment of criminal recidivism because it does not require detection or prosecution by law enforcement. Because many crimes are unreported by victims and undetected by the authorities (Truman & Langton, 2014), arrest and conviction data may underestimate true levels of criminal activity. For obvious reasons, however, individuals cannot be relied upon to acknowledge their crimes unless they receive strict assurances that the information will be kept confidential and will not be used against them in a criminal proceeding. Drug Courts will typically be required to hire an independent evaluator who has no connection to the court or criminal justice system to confidentially survey participants. This method is likely to be prohibitively costly for many Drug Courts, which explains why it has rarely been employed with the notable exception of one highly funded national study (Rossman et al., 2011).

Whether measured by arrests, convictions, or incarcerations, categorizing recidivism according to the level (i.e., felony, misdemeanor, or summary offense) and nature (e.g., drug offenses, property and theft offenses, violent offenses, technical violations, prostitution, and traffic offenses) of the crimes involved is highly informative and necessary. Different categories of crime can have very different implications for public safety and cost. For example, violent offenses may have serious victimization costs and may result in substantial jail or prison sentences, whereas drug possession may not involve an identifiable victim and is more likely to receive a less costly probation sentence (Zarkin et al., 2012).

As a final note, not all Drug Courts have reasonable access to data on new arrests, convictions, or incarcerations occurring after participants have been discharged from the program. In some jurisdictions, these records may be in the possession of other executive agencies, such as the police department or department of corrections, and the Drug Court may not be entitled to the information. Under such circumstances, Drug Courts should make every effort to negotiate access to the data, but of course, Drug Courts cannot be held accountable for reporting information beyond their reach.

D. Independent Evaluations

In addition to monitoring their own performance, Drug Courts benefit greatly from having an independent evaluator examine their program and issue recommendations to improve their adherence to best practices. Drug Courts that engaged an independent evaluator and implemented at least some of the evaluator's recommendations were determined in one multisite study to be twice as cost-effective and nearly twice as effective at reducing crime as Drug Courts that did not engage an independent evaluator (Carey et al., 2008, 2012).

Drug Courts benefit from an independent evaluation for several reasons. Every program has blind spots that prevent staff from recognizing their own shortcomings. Some team members, such as the judge, may have more social influence or power than others, making it difficult for some team members to call attention to problems in court or during team meetings. Drug Courts also operate in a political environment and staff may

be hesitant to criticize local practices for fear of reprisal. An independent evaluator from another jurisdiction can usually offer frank criticisms of current practices with less fear of repercussions (Heck & Thanner, 2006).

Although most Drug Courts are capable of keeping descriptive statistics about their program, considerably more expertise is required to perform *inferential analyses*, which compare Drug Court outcomes to those of a comparison group. Controlling statistically for preexisting group differences that could bias one's results is often necessary. For example, if Drug Court participants had fewer previous convictions than comparison subjects before entering the study, better outcomes for the Drug Court might simply reflect the fact that it treated a less severe population. Evaluators must take numerous scientific matters into consideration and may need to apply several levels of statistical corrections to produce valid and reliable results.

Studies also reveal that participant perceptions are often highly predictive of outcomes in Drug Courts. For example, perceptions concerning the procedural fairness of the program (Burke, 2010; McIvor, 2009), the manner in which incentives and sanctions are delivered (Goldkamp et al., 2002; Harrell & Roman, 2001; Marlowe et al., 2005), and the quality of the treatment services provided (Turner et al., 1999) are often predictive of recidivism and correlate significantly with adherence to best practices. Needless to say, participants are more likely to be forthright with an independent evaluator about their perceptions of the Drug Court than with staff members who control their fate in the criminal justice system.

Studies have not determined how frequently Drug Courts should be evaluated by an independent investigator. Generally speaking, a new evaluation should be performed whenever a program or the environment within which it operates changes substantially. Staff turnover and evidence of drift from the intended model are critical events that call for a new evaluation (Yeaton & Camberg, 1997). Evidence suggests that staff turnover and model drift occur within five-year intervals in Drug Courts. Within five years, between roughly 30% and 60% of Drug Courts experience substantial turnover in key staff positions (Van Wormer, 2010). The highest turnover rates, commonly exceeding 50%, are among substance use disorder and mental health treatment providers (Lutze & Van Wormer, 2007; McLellan et al., 2003; Taxman & Bouffard, 2003; Van Wormer, 2010). Evidence further reveals that staff turnover correlates significantly with drift in the quality of the services provided (Van Wormer, 2010). Therefore, five years is a reasonable outside estimate of how frequently Drug Courts should be evaluated independently. If resources allow, Drug Courts should engage independent evaluators at more frequent intervals to detect drift readily and prevent services from worsening with time.

Drug Courts need to select competent evaluators. The first step in selecting a competent evaluator is to request recommendations from other Drug Courts and national organizations that are familiar with Drug Court operations and research. Senior staff at NADCP and NDCI are familiar with the evaluation literature and the skill sets of dozens of evaluators nationally. When selecting an evaluator, review prior evaluation reports, especially those involving Drug Courts or other problem-solving courts. If prior evaluations failed to follow the practices described herein, consider selecting another evaluator who has demonstrated expertise in applying best practices related to Drug Court program evaluations. One of the most important questions to consider when reviewing prior evaluations is whether the report recommended concrete actions the Drug Court could take to enhance its adherence to best practices and improve its outcomes. The most effective evaluators are aware of the literature on best practices, measure Drug Court practices against established performance benchmarks, and promote useful strategies to improve each program's operations and results.

Many Drug Courts do not have sufficient resources to hire independent evaluators. One way to address this problem is to contact local colleges or universities to determine whether graduate or undergraduate students may be interested in evaluating the Drug Court as part of a thesis, dissertation, or capstone project. Because such projects require close supervision from senior academic faculty, the Drug Court can receive high-level research expertise at minimal or no cost. Moreover, students are likely to be highly motivated to complete the evaluation successfully because their academic degree and standing depends on it.

E. Historically Discriminated Against Groups

The focus of this section in on socio-demographic groups that have historically experienced sustained discrimination or reduced social opportunities due to their race, ethnicity, gender, sexual orientation, sexual identity, physical or mental disability, religion, or socioeconomic status. Best practices for ensuring

equivalent treatment for members of groups that have historically experienced discrimination in Drug Courts are described in Standard II, Equity and Inclusion.

Evidence suggests racial and ethnic minority individuals are underrepresented in some Drug Courts and may have lower graduation rates than other participants [see Commentary in Standard II, Equity and Inclusion]. Drug Courts have an affirmative obligation to determine whether racial and ethnic minority individuals and members of other groups that have historically experienced discrimination are being disproportionately burdened or excluded from their programs; and if so, to take reasonable corrective measures to rectify the problem and evaluate the success of the corrective actions [see Standard II]. Not knowing whether one's Drug Court is disproportionately burdening disadvantaged groups is itself a violation of best practice standards (Marlowe, 2013).

Studies have not determined how frequently Drug Courts should review performance information for members of groups that have historically experienced discrimination. Consistent with the general literature on CPI, CQI and MFR, the Drug Court team should review performance information at least annually and implement and evaluate self-corrective measures on a rapid-cycle basis (Rudes et al., 2013; Wexler et al., 2012).

A number of resources are available to help Drug Courts identify and rectify disparate impacts for groups that have historically experienced discrimination (e.g., Casey et al., 2012; Rubio et al., 2008b; Yu et al., 2009). Seasoned evaluators and university faculty are likely to be familiar with this literature and to know how to perform these types of analyses. Many analyses, such as comparing graduation rates between different racial groups, are relatively simple and straightforward to perform. Other analyses, such as determining whether disadvantaged groups have equivalent access to Drug Court, are considerably more difficult. Many Drug Courts may not have adequate information about the relevant arrestee population to determine whether disadvantaged groups are gaining access to the Drug Court at equivalent rates. For example, information might not be available to determine what proportion of racial-minority arrestees have serious drug problems and are therefore eligible for participation in Drug Court. The primary challenge for such Drug Courts may be to gain better access to a wider range of information on the arrestee population, and as a practical matter, such analyses may be beyond the ability and expertise of some programs to accomplish.

F. Electronic Database

Paper files have minimal value for conducting program evaluations. Evaluators are typically required to extract information from handwritten notes and progress reports that are difficult to read, contain contradictory information, and have numerous missing entries. As a consequence, many evaluations are completed months or years after the fact when the results may no longer reflect what is occurring in the program. Such evaluations often contain so many gaps or caveats in the data that the conclusions which may be drawn are tentative at best.

Drug Courts are approximately 65% more cost-effective when they enter standardized information concerning their services and outcomes into an electronic management information system (MIS), which is capable of generating automated summary reports (Carey et al., 2008, 2012). The cost of purchasing an MIS is offset many times over by providing greater efficiencies in operations and yielding the type of performance feedback that is necessary to continually improve and fine-tune one's Drug Court program.

Appendix E provides examples of MISs that have been developed for use in Drug Court evaluations. Some of the older and less sophisticated systems can be obtained free of charge. For example, the Buffalo System (so named because it was developed in a Drug Court in Buffalo, New York) is a Microsoft Access database that can be obtained at no cost by contacting NADCP. Newer systems must be purchased or licensed, but are more likely to be web-based and can be accessed simultaneously by multiple users and agencies. Allowing multiple agencies to use the same MIS, each with its own secured and encrypted access, can spread the cost of the system across several budgets. Newer systems are also more likely to have preprogrammed analytic reports that provide important summary information for staff at the push of a button. Finally, newer systems are more likely to include a data-extraction tool. A data-extraction tool allows information to be imported readily into a statistical program, such as SAS or SPSS, which skilled evaluators then can use to conduct sophisticated statistical analyses.

G. Timely and Reliable Data Entry

The biggest threat to a valid program evaluation is poor data entry by staff. The adage "garbage in/garbage out" is particularly apt in this regard. If staff members do not accurately record what occurred, no amount of scientific expertise or sophisticated statistical adjustments can produce valid findings.

The best time to record information about services and events is when they occur. For example, staff members should enter attendance information into an MIS or written log during court hearings and treatment sessions. This is referred to as *real-time recording*. The typical staff person in a Drug Court is responsible for dozens of participants and each participant has multiple obligations in the program, such as appearing at court hearings, attending treatment sessions, and delivering urine specimens. Only the rare staff person can recall accurately what events transpired or should have transpired days or weeks in the past. Attempting to reconstruct events from memory is likely to introduce unacceptable error into a program evaluation.

Data should ordinarily be recorded within no more than forty-eight hours of the respective events. Medicare, for instance, requires physicians to document services within a "reasonable time frame," defined as twenty-four to forty-eight hours (Pelaia, n.d.). After forty-eight hours, errors in data entry have been shown to increase significantly. After one week, information is so likely to be inaccurate that it may be better to leave the data as missing than attempt to fill in gaps from faulty memory (Marlowe, 2010).

Staff members who are persistently tardy when entering data pose a serious threat to the integrity of a Drug Court. Not only are evaluation results unlikely to be accurate, but those same staff persons are unlikely to be delivering appropriate services. Good-quality treatment and supervision require staff to monitor participant behavior vigilantly, record performance information in a timely and actionable fashion, and adjust services and consequences accordingly. Failing to record performance information in a timely and reliable manner undermines the quality and effectiveness of a Drug Court and seriously jeopardizes participant care.

H. Intent-to-Treat Analyses

A serious error in some Drug Court evaluations is to examine outcomes only for participants who graduated successfully from the program. The logic for performing such an analysis is understandable. Evaluators are often interested in learning what happens to individuals who received all of the services the program has to offer. If individuals who dropped out or were terminated prematurely from the Drug Court are included in the analyses, the results will be influenced by persons who did not receive all of the intended services.

Although this reasoning might seem logical, it is scientifically flawed (Heck, 2006; Heck & Roussell, 2007; Marlowe, 2010, in press; Peters, 1996; Rempel, 2006, 2007). Outcomes must be examined for all eligible individuals who participated in the Drug Court regardless of whether they graduated, were terminated, or withdrew from the program. This is referred to as an *intent-to-treat analysis* because it examines outcomes for all individuals whom the program initially set out to treat. Reporting outcomes for graduates alone is not appropriate because such an analysis unfairly and falsely inflates the apparent success of the program. For example, individuals who graduated from the Drug Court are more likely than terminated participants to have entered the program with less severe drug or alcohol problems, less severe criminal propensities, higher motivation for change, or better social supports. As a result, they might have been less likely to commit future offenses or relapse to substance use regardless of the services they received in Drug Court.

This issue is particularly important when outcomes are contrasted against those of a comparison sample, such as probationers. Selecting the most successful Drug Court cases and comparing their outcomes to all of the probationers unfairly skews the results in favor of the Drug Court. It is akin to selecting the A+ students from one classroom, comparing their scores on a test to those of all of the students in a second classroom, and then concluding the first class had a better teacher. Such a comparison would clearly be slanted unfairly in favor of the first teacher.

This is not to suggest that outcomes for graduates are of no interest. Drug Courts may, indeed, want to know what happens to individuals who receive all of the services in the program. This, however, should be a *secondary analysis* that is performed after the intent-to-treat analysis has shown positive results. If it is first determined that the Drug Court achieved significantly improved outcomes on an intent-to-treat basis, it may

then be appropriate to proceed further and determine whether outcomes were even better for the graduates. If the intent-to-treat analysis is not significant, then it is not acceptable to move on to evaluate outcomes for graduates alone.

Importantly, if secondary analyses are performed on Drug Court graduates, then the comparison sample should also comprise successful completers. For example, outcomes for Drug Court graduates should be compared to those of probationers who satisfied the conditions of probation. Comparing outcomes for Drug Court graduates to all probationers, including probation failures, would unfairly favor the Drug Court.

The only exception to an intent-to-treat analysis is for what are sometimes referred to as *neutral discharges*. Some Drug Courts assign a neutral discharge to participants who are withdrawn from the program for reasons beyond the control of the participant and the program. A neutral discharge is assigned most commonly when the Drug Court discovers a participant was admitted to the program erroneously. For example, a participant might need to be withdrawn from Drug Court if he or she had a prior conviction that precluded eligibility for the Drug Court or resided in a judicial district that was not within the jurisdictional boundaries of the Drug Court. A neutral discharge may also be assigned to participants who are withdrawn from the program because they enlisted in the military or moved out of the jurisdiction with the court's permission. A neutral discharge should never be assigned to cases in which termination was related to a participant's performance in Drug Court.

I. Comparison Groups

The mere fact that individuals perform well after participating in Drug Court does not prove the Drug Court was responsible for their favorable outcomes. Those same individuals might have functioned just as well if they had never entered Drug Court. To examine the important question of causality, the performance of Drug Court participants must be compared against that of an equivalent and unbiased comparison group. Comparing what happened in the Drug Court to what would most likely have happened if the Drug Court did not exist is referred to as testing the *counterfactual hypothesis*, or the possibility that the Drug Court was ineffective (Popper, 1959).

Some comparison groups are reasonably unbiased and can yield a fair and accurate assessment of what would most likely have occurred without the Drug Court. Others, however, may be systematically biased in such a manner as to make the Drug Court look better or worse than it deserves. This may lead to the unwarranted conclusion that the Drug Court was effective or ineffective when, in fact, the reverse could be true.

Random Assignment—The strongest inference of causality may be reached when eligible individuals are randomly assigned either to the Drug Court or to a comparison group. Random assignment provides the greatest assurance that the groups started out with an equal chance of success; therefore, better outcomes for one group can be confidently attributed to the effects of the program (Campbell & Stanley, 1963; Farrington, 2003; Farrington & Welsh, 2005; National Research Council, 2001; Telep et al., 2015). Even when an evaluator employs random assignment, there is still the possibility (albeit a greatly diminished possibility) that the groups differed on important dimensions from the outset. This possibility requires the evaluator to perform a confirmation of the randomization procedure. The evaluator will need to check for preexisting differences between the groups that could have affected the results. If the groups differed significantly on variables that are correlated with outcomes (such as the severity of participants' criminal histories or drug problems), the evaluator might employ statistical procedures to adjust for those differences and obtain defensible results.

As a practical matter, conducting random assignment is often very difficult in Drug Courts. Some staff members may have ethical objections against denying potentially effective services to eligible individuals. Moreover, some Drug Courts may have difficulty filling their slots and may not wish to turn away eligible individuals. The evaluator will also need to gain approval and buy-in for random assignment from numerous professionals and agencies, including the court, prosecution, and defense counsel. Finally, random assignment usually requires implementation of ethical safeguards (National Research Council, 2001). For example, participants may need to provide informed consent to random assignment, and an independent ethics review board may need to oversee the safety and fairness of the study. Local colleges and universities

often have institutional review boards (IRBs) or data and safety monitoring boards (DSMBs) which have the authority and expertise to provide ethical oversight for randomized studies.

Random assignment poses far fewer challenges if a Drug Court has insufficient capacity to treat many individuals who would otherwise be eligible for its services. If many eligible people must be turned away, then it would arguably be fairest to select participants randomly rather than allow staff members to pick and choose who gets into the program. Under such circumstances, random assignment may provide the best protection against unfair discrimination and unconscious bias (National Research Council, 2001). In fact, a number of Drug Court studies have used random assignment successfully in light of insufficient program capacity (e.g., Gottfredson et al., 2003; Jones, 2011; Turner et al., 1999).

Quasi-Experimental Comparison Group—In many Drug Courts, engaging in random assignment is simply impractical. The next best approach is to use a quasi-experimental comparison group (Campbell & Stanley, 1963). This refers to individuals who were eligible for the Drug Court but did not enter for reasons that are unlikely to have influenced their outcomes. Perhaps the best example is individuals who were eligible for and willing to enter the Drug Court, but were denied access because there were no empty slots available. This is referred to as a *wait-list comparison group*. The mere happenstance that the Drug Court was full is unlikely to have led to the systematic exclusion of individuals who had more severe problems or poorer prognoses to begin with, and therefore is unlikely to bias the results.

Less optimal, but still potentially acceptable, quasi-experimental comparison groups include individuals who would have been eligible for the Drug Court but were arrested in the year or so before the Drug Court was established, or were arrested in an immediately adjacent county that does not have a Drug Court (Heck, 2006; Heck & Roussell, 2007; Marlowe, 2010, in press; Peters, 1996). Because these individuals were arrested at an earlier point in time or in a different geographic region than the Drug Court participants, such comparison groups might still be different enough from the Drug Court group to bias the results. For example, socioeconomic conditions might differ significantly between neighboring communities, or law enforcement practices might change from year to year. The likelihood of this occurring, however, is usually not substantial and these may be the only practical comparison conditions that can be used for many Drug Court evaluations.

When using a quasi-experimental comparison group, the evaluator must check for preexisting differences between the groups that could have affected the results (Campbell & Stanley, 1963). For example, the comparison individuals may have had more serious criminal histories than the Drug Court participants to begin with. This, in turn, might have put them at greater risk for criminal recidivism. If so, then superior outcomes for the Drug Court participants might not have been due to the effects of the Drug Court, but rather to the fact that it treated a less severe population. A skilled evaluator can use a number of statistical procedures to adjust for such differences and potentially obtain scientifically defensible results.

Matched Comparison Group—Evaluators do not always have a quasi-experimental comparison group at their disposal. Under such circumstances, they may be required to construct a comparison group out of a large and heterogeneous pool of offenders. For example, an evaluator might need to select comparison subjects from a statewide probation database. Many of those probationers would not have been eligible for Drug Court, or are dissimilar to Drug Court participants on characteristics that are likely to have influenced their outcomes. For example, some of the probationers might not have had serious drug problems, or might have been charged with offenses that would have excluded them from participation in Drug Court. The evaluator must, therefore, select a subset of individuals from the entire probation pool that are similar to the Drug Court participants on characteristics that are known to affect outcomes. For example, the evaluator might pair each Drug Court participant with a probationer who has the same or similar criminal history, demographic characteristics, and substance use diagnosis (Heck, 2006; Marlowe, 2010, in press). Because the evaluator will choose only those probationers who are similar to the Drug Court participants on multiple characteristics, it is necessary to start out with a large sample of potential candidates from which to select comparable individuals.

The success of any matching strategy will depend largely on whether the evaluator has adequate information about the comparison candidates to make valid matches (Campbell & Stanley, 1963). If data are not available on such important variables as the probationers' criminal histories or substance use problems, evaluators and Drug Courts will not be able to place confidence in the validity of the matches. Simply matching the groups

on variables that are easy to measure and readily available, such as gender or race, is not sufficient because the groups might differ on other important dimensions that were not taken into account.

Propensity Score Analysis—An evaluator may also use an advanced statistical procedure called a propensity score analysis to mathematically adjust for differences between the Drug Court and comparison groups. This procedure calculates the statistical probability that an individual with a given set of characteristics would be in the Drug Court group as opposed to the comparison group—in other words, the relative similarity of that individual to one group as opposed to the other (Dehejia & Wahba, 2002). The analysis then mathematically adjusts for this relative similarity when comparing outcomes. Advanced statistical expertise is required to implement and interpret this complicated procedure.

As with any statistical adjustment, the success of a propensity score analysis will depend on whether the evaluator has adequate information about the comparison subjects to make valid adjustments. If data are not available on such important variables as the comparison subjects' criminal histories or substance use problems, evaluators and Drug Courts will not be able to place confidence in the adjustments (Peikes et al., 2008). Again, merely adjusting the scores based on easily measured variables, such as gender or race, is not sufficient because the groups might differ on other important dimensions that were never taken into account.

Invalid Comparison Groups—Several comparison groups have been used in Drug Court evaluations that quite likely produced seriously biased results. Comparing outcomes from a Drug Court to those of individuals who refused to enter the Drug Court, were denied access to the Drug Court because of their clinical or criminal histories, dropped out of the Drug Court, or were terminated prematurely from the Drug Court is rarely, if ever, justified (Heck, 2006; Heck & Thanner, 2006; Marlowe, 2010, in press; Peters, 1996). The probability is unacceptably high that such persons had poorer prognoses or more severe problems to begin with. For example, they very likely had more serious criminal or substance use histories, lower motivation for change, or lesser social supports. Given the high likelihood that these individuals were seriously disadvantaged from the outset, statistical adjustments cannot be relied upon to overcome the differences (Campbell & Stanley, 1963).

J. Time at Risk

For an evaluation to be valid, Drug Court and comparison participants must have the same time at risk, meaning the same opportunity to engage in substance use, crime, and other behaviors of interest to the evaluation. If, for example, an evaluator measured criminal recidivism over a period of twelve months for Drug Court participants, but over a period of twenty-four months for the comparison group, this would give an unfair advantage to the Drug Court participants. The comparison group participants would have twelve additional months in which to commit new crimes or other infractions.

Ensuring an equivalent time at risk requires the evaluator to begin the analyses from a comparable start date for both groups. As was mentioned earlier, Drug Court evaluations typically use the date of entry into Drug Court or the date of the arrest or technical violation that made the individual eligible for Drug Court as the start date for analyses. If the comparison group is comprised of probationers, comparable start dates might be the date the individual was placed on probation or the date of the arrest that led to a probation sentence.

If the time at risk differs significantly between groups, the evaluator might be able to compensate for this problem by adjusting statistically for time at risk in outcome comparisons. For example, the evaluator might enter time at risk as a covariate in the statistical analyses. A *covariate* is a variable that is entered first into a statistical model. The independent effect of the variable of interest (in this case, being treated in a Drug Court) is then examined after first taking the effect of the covariate into account. This procedure would indicate whether Drug Court participants had better outcomes after first taking into account the influence of their shorter time at risk. The use of covariates is not always successful, however, and the best course of action is to ensure the groups have equivalent follow-up windows.

A related issue is referred to as *time at liberty*. Time at liberty and time at risk are similar in that both affect a participant's opportunity to reoffend or engage in other behaviors of interest to the evaluation. The difference is that time at liberty relates to whether restrictive conditions were placed on the participant. The most obvious restrictive conditions involve physical barriers to freedom, such as incarceration or placement

in a residential treatment facility. These physical barriers severely restrict a participant's ability to use drugs, commit new offenses, obtain a job, or engage in other behaviors of interest to evaluators.

A potential error in Drug Court evaluations is to neglect time at liberty when performing outcome comparisons. In some jurisdictions, for example, individuals who do not enter Drug Court may be more likely to receive a jail sentence. If they are jailed for a portion of the follow-up period, they might have fewer opportunities to reoffend or use drugs than Drug Court participants who are treated in the community. The evaluator might conclude, erroneously, that Drug Court caused participants to reoffend or use drugs more often, when in fact they simply had more time at liberty to do so. Under such circumstances, the evaluator would need to adjust statistically for participants' time at liberty in the outcome analyses. For example, the evaluator might need to enter time at liberty as a covariate in the statistical models. This would indicate whether Drug Court participants had better outcomes after first taking into account their longer time at liberty. As was noted earlier, such adjustments are not always successful and Drug Courts will require expert consultation to ensure the analyses are carried out appropriately.

Note that evaluators are not always advised to adjust for time at liberty. In cost analyses, for example, the time participants spend in jail or a residential treatment facility is an important outcome in its own right and should be valued accordingly from a fiscal standpoint. Deciding whether to adjust for time at liberty, like many evaluation-related decisions, requires scientific expertise and careful consideration of the aims of the study. For such analyses, Drug Courts are strongly advised to obtain expert statistical and scientific consultation.

REFERENCES

Aos, S., Miller, M., & Drake, E. (2006). *Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates*. Olympia: Washington State Institute for Public Policy.

Belenko, S., Patapis, N., & French, M.T. (2005). *Economic benefits of drug treatment: A critical review of the evidence for policy makers.* Philadelphia: Treatment Research Institute.

Berman, G., Rempel, M., & Wolf, R.V. (Eds.). (2007). *Documenting results: Research on problem-solving justice.* New York: Center for Court Innovation.

Bryson, J.M., Crosby, B.C., & Stone, M.M. (2006). The design and implementation of cross-sector collaborations: Propositions from the literature. *Public Administration Review, 66*(Suppl.1), 44–55.

Burke, K.S. (2010). Just what made drug courts successful? *New England Journal on Criminal & Civil Confinement, 36*(1), 39–58.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally College Publishing Company.

Carey, S.M., Finigan, M.W., & Pukstas, K. (2008). *Exploring the key components of drug courts: A comparative study of 18 adult drug courts on practices, outcomes and costs.* Portland, OR: NPC Research.

Carey, S.M., Mackin, J.R., & Finigan, M.W. (2012). What works? The ten key components of drug court: Research-based best practices. *Drug Court Review, 8*(1), 6–42.

Carey, S.M., & Waller, M.S. (2011). *Oregon drug courts: Statewide costs and promising practices.* Portland, OR: NPC Research.

Casey, P., Warren, R., Cheesman, F., & Elek, J. (2012). *Helping courts address implicit bias: Resources for education.* Williamsburg, VA: National Center for State Courts.

Cissner, A., Rempel, M., Franklin, A.W., Roman, J.K., Bieler, S., Cohen, R., & Cadoret, C.R. (2013). *A statewide evaluation of New York's adult drug courts: Identifying which policies work best.* New York: Center for Court Innovation.

Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S.R., Alexander, J.A., & Lowery, J.C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*(50). doi:10.1186/1748-5908-4-50

Dehejia, R.H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*(1), 151–161.

DeVall, K.E., Gregory, P.D., & Hartmann, D.J. (2015, June 10). Extending recidivism monitoring for drug courts: Methods, issues and policy implications. *International Journal of Offender Therapy and Comparative Criminology: Online.* doi:10.1177/0306624X15590205.

Downey, P.M., & Roman, J.K. (2010). *A Bayesian meta-analysis of drug court cost-effectiveness.* Washington, DC: The Urban Institute.

Durose, M.R., Cooper, A.D., & Snyder, H.N. (2014). *Recidivism of prisoners released in 30 states in 2005: Patterns from 2005 to 2010.* Washington, DC: U.S. Dept. of Justice, Bureau of Justice Statistics.

Etheridge, R.M., Craddock, S.G., Dunteman, G.H., & Hubbard, R.L. (1995). Treatment services in two national studies of community-based drug abuse treatment programs. *Journal of Substance Abuse, 7*(1), 9–26.

Farrington, D.P. (2003). A short history of randomized experiments in criminology: A meagre feast. *Evaluation Review, 27*(3), 218–227.

Farrington, D.P., & Welsh, B.C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology, 1*(1), 9–38.

Friedmann, P.D., Taxman, F.S., & Henderson, C.E. (2007). Evidence-based treatment practices for drug-involved adults in the criminal justice system. *Journal of Substance Abuse Treatment, 32*(3), 267–277.

Gifford, E.J., Eldred, L.M., McCutchan, S.A., & Sloan, F.A. (2014). The effects of participation level on recidivism: A study of Drug Treatment Courts using propensity score matching. *Substance Abuse Treatment, Prevention, and Policy, 9*(40). doi:10.1186/1747-597X-9-40

Goldkamp, J.S., White, M.D., & Robinson, J.B. (2002). An honest chance: Perspectives on drug courts. *Federal Sentencing Reporter, 14*(6), 369–372.

Gossop, M., Tradaka, K., Stewart, D., & Witton, J. (2005). Reductions in criminal convictions after addiction treatment: 5-year follow-up. *Drug & Alcohol Dependence, 79*(3), 295–302.

Gottfredson, D.C., Kearley, B.W., & Bushway, S.D. (2008). Substance use, drug treatment, and crime: An examination of intra-individual variation in a drug court population. *Journal of Drug Issues, 38*(2), 601–630.

Gottfredson, D.C., Kearley, B.W., Najaka, S.S., & Rocha, C.M. (2007). How Drug Treatment Courts work: An analysis of mediators. *Journal of Research on Crime & Delinquency, 44*(1), 3–35.

Gottfredson, D.C., Najaka, S.S., & Kearley, B. (2003). Effectiveness of Drug Treatment Courts: Evidence from a randomized trial. *Criminology & Public Policy, 2*(2), 171–196.

Government Accountability Office. (2011). *Adult drug courts: Studies show courts reduce recidivism, but DOJ could enhance future performance measure revision efforts* [No. GAO-12-53]. Washington, DC: Author.

Harrell, A., & Roman, J. (2001). Reducing drug use and crime among offenders: The impact of graduated sanctions. *Journal of Drug Issues, 31*(1), 207–232.

Heck, C. (2006). *Local drug court research: Navigating performance measures and process evaluations* [Monograph Series No. 6]. Alexandria, VA: National Drug Court Institute. *Available at* http://www.ndci.org/sites/default/files/ndci/Mono6.Local Research.pdf

Heck, C., & Roussell, A. (2007). Record keeping and statistics. In J.E. Lessenger & G.F. Roper (Eds.), *Drug courts: A new approach to treatment and rehabilitation* (pp. 401–413). New York: Springer.

Heck, C., & Thanner, M.H. (2006). Evaluating drug courts: A model for process evaluation. *Drug Court Review, 5*(2), 51–82.

Henderson, C.E., Young, D.W., Jainchill, N., Hawke, J., Farkas, S., & Davis, R.M. (2007). Program use of effective drug abuse treatment practices for juvenile offenders. *Journal of Substance Abuse Treatment, 32*(3), 279–290.

Huebner, B.M., & Cobbina, J. (2007). The effect of drug use, drug treatment participation, and treatment completion on probationer recidivism. *Journal of Drug Issues, 37*(3), 619–641.

Inciardi, J., Martin, S., & Butzin, C. (2004). Five-year outcomes of therapeutic community treatment of drug-involved offenders after release from prison. *Crime & Delinquency, 50*(1), 88–107.

Jones, C. (2011, November). Intensive judicial supervision and drug court outcomes: Interim findings from a randomized controlled trial. *Crime & Justice Bulletin, 152,* 1–16. *Available at* http://www.bocsar.nsw.gov.au/Documents/cjb152.pdf

Jones, C., & Kemp, R.I. (2011). The relationship between early-phase substance use trajectories and drug court outcomes. *Criminal Justice & Behavior, 38*(9), 913–933.

King, R., & Elderbroom, B. (2014). *Improving recidivism as a performance measure.* Washington, DC: The Urban Institute. *Available at* http://www.urban.org/UploadedPDF/413247-improving-recidivism.pdf

Knight, K., Simpson, D.D., & Hiller, M.L. (1999). Three-year reincarceration rates outcomes for in-prison therapeutic community treatment in Texas. *Prison Journal, 79*(3), 337–351.

Latimer, J., Morton-Bourgon, K., & Chretien, J. (2006). *A meta-analytic examination of drug treatment courts: Do they reduce recidivism?* Ottawa, ON: Canada Dept. of Justice, Research & Statistics Division.

Lowenkamp, C.T., Holsinger, A.M., & Latessa, E.J. (2005). Are drug courts effective? A meta-analytic review. *Journal of Community Corrections, 15*(1), 5–28.

Lutze, F.E., & Van Wormer, J.G. (2007). The nexus between drug and alcohol treatment program integrity and drug court effectiveness: Policy recommendations for pursuing success. *Criminal Justice Policy Review, 18*(3), 226–245.

Marlowe, D.B. (2010). *Introductory handbook for DWI court program evaluations.* Alexandria, VA: National Center for DWI Courts. *Available at* http://www.dwicourts.org/sites/default/files/nadcp/DWI%20Ct%20Eval%20Manual%20REVISED-8-10.pdf

Marlowe, D.B. (2013). Achieving racial and ethnic fairness in drug courts. *Court Review, 49*(1), 40–47.

Marlowe, D.B. (in press). *Manual for scientific monitoring and evaluation of drug treatment courts in the Americas.* Washington, DC: Inter-American Drug Abuse Control Commission, Organization of American States.

Marlowe, D.B., Festinger, D.S., Foltz, C., Lee, P.A., & Patapis, N.S. (2005). Perceived deterrence and outcomes in drug court. *Behavioral Sciences & the Law, 23*(2), 183–198.

Marlowe, D.B., Heck, C., Huddleston, C.W., & Casebolt, R. (2006). A national research agenda for drug courts: Plotting the course for second-generation scientific inquiry. *Drug Court Review, 5*(2), 1–31.

Martin, S.S., Butzin, C.A., Saum, C.A., & Inciardi, J.A. (1999). Three-year reincarceration outcomes of therapeutic community treatment for drug-involved offenders in Delaware: From prison to work release to aftercare. *Prison Journal, 79*(3), 294–320.

McIvor, G. (2009). Therapeutic jurisprudence and procedural justice in Scottish drug courts. *Criminology & Criminal Justice, 9*(1) 29–49.

McLellan, A.T., Carise, D., & Kleber, H.D. (2003). Can the national addiction treatment infrastructure support the public's demand for quality care? *Journal of Substance Abuse Treatment, 25*(2), 117–121.

Mitchell, O., Wilson, D.B., Eggers, A., & MacKenzie, D.L. (2012). Assessing the effectiveness of drug courts on recidivism: A meta-analytic review of traditional and nontraditional drug courts. *Journal of Criminal Justice, 40*(1), 60–71.

National Institute of Justice. (2010). Drug court performance measures and program evaluation. Retrieved from http://www.nij.gov/topics/courts/drug-courts/pages/measures-evaluation.aspx

National Research Council. (2001). *Informing America's policy on illegal drugs: What we don't know keeps hurting us.* Washington, DC: National Academy Press.

Olson, D.E., & Lurigio, A.J. (2014). The long-term effects of prison-based drug treatment and aftercare services on recidivism. *Journal of Offender Rehabilitation, 53*(8), 600–619.

Peikes, D.N., Moreno, L., & Orzol, S.M. (2008). Propensity score matching: A note of caution for evaluators of social programs. *American Statistician, 62*(3), 222–231. doi:10.1198/000313008X332016

Pelaia, R.A. (n.d.). Medical record entries: What is timely and reasonable? Retrieved from http://news.aapc.com/index.php/2013/09/medical-record-entries-what-is-timely-and-reasonable/

Peters, R.H. (1996). *Evaluating drug court programs: An overview of issues and alternative strategies.* Washington, DC: Justice Programs Office at American University.

Peters, R.H., Haas, A.L., & Hunt, W.M. (2002). Treatment "dosage" effects in drug court programs. *Journal of Offender Rehabilitation, 33*(4), 63–72.

Popper, K. (1959). *The logic of scientific discovery.* New York: Harper & Row.

Rempel, M. (2006). Recidivism 101: Evaluating the impact of your drug court. *Drug Court Review, 5*(2), 83–112.

Rempel, M. (2007). Action research: Using information to improve your drug court. In G. Berman, M. Rempel & R.V. Wolf (Eds.), *Documenting results: Research on problem-solving justice* (pp. 101–122). New York: Center for Court Innovation.

Rossman, S.B., Rempel, M., Roman, J.K., Zweig, J.M., Lindquist, C.H., Green, M.,… Farole, D.J. (2011). *The multisite adult drug court evaluation: The impact of drug courts (vol. 4).* Washington, DC: Urban Institute Justice Policy Center. Available at https://www.ncjrs.gov/pdffiles1/nij/grants/237112.pdf

Rubio, D.M., Cheesman, F., & Federspiel, W. (2008a). *Performance measurement of drug courts: The state of the state.* Williamsburg, VA: National Center for State Courts. *Available at* http://cdm16501.contentdm.oclc.org/cdm/ref/collection/spcts/id/171

Rubio, D.M., Cheesman, F., & Webster, L. (2008b). *Kentucky drug court statewide technical assistance project: Development of statewide adult drug court performance measures*. Denver, CO: National Center for State Courts.

Rudes, D.S., Viglione, J., & Porter, C.M. (2013). Using quality improvement models in correctional organizations. *Federal Probation, 77*(2). *Available at* http://www.uscourts.gov/uscourts/FederalCourts/PPS/FedProb/2013-09/quality.html

Shaffer, D.K. (2010). Looking inside the black box of drug courts: A meta-analytic review. *Justice Quarterly, 28*(3), 493–521.

Taxman, F.S., & Belenko, S. (2013). *Implementing evidence-based practices in community corrections and addiction treatment*. New York: Springer.

Taxman, F.S., & Bouffard, J.A. (2003). Substance abuse counselors' treatment philosophy and the content of treatment services provided to offenders in drug court programs. *Journal of Substance Abuse Treatment, 25*(2), 75–84.

Taxman, F.S., Perdoni, M.L., & Harrison, L.D. (2007). Drug treatment services for adult offenders: The state of the state. *Journal of Substance Abuse Treatment, 32*(3), 239–254.

Telep, C.W., Garner, J.H., & Visher, C.A. (2015, July 3). The production of criminological experiments revisited: The nature and extent of federal support for experimental designs, 2001–2013. *Journal of Experimental Criminology: Online.* doi:10.1007/s11292-015-9239-6

Truman, J.L., & Langton, L. (2014). *Criminal victimization, 2013.* Washington, DC: Bureau of Justice Statistics, U.S. Dept. of Justice. *Available at* http://www.bjs.gov/content/pub/pdf/cv13.pdf

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

Van Wormer, J. (2010). Understanding operational dynamics of drug courts (Doctoral dissertation, University of Washington). Retrieved from http://research.wsulibs.wsu.edu:8080/xmlui/bitstream/handle/2376/2810/vanWormer_wsu_0251E_10046.pdf ?sequence=1

Waters, N.L., Cheesman, F.L., Gibson, S.A., & Dazevedo, I. (2010). *Mental health court performance measures: Implementation and user's guide*. Williamsburg, VA: National Center for State Courts.

Wexler, H.K., Melnick, G., Lowe, L., & Peters, J. (1999). Three-year reincarceration outcomes for Amity in-prison therapeutic community and aftercare in California. *Prison Journal, 79*(3), 321–336.

Wexler, H.K., Zehner, M., & Melnick, G. (2012). Improving drug court operations: NIATx organizational improvement model. *Drug Court Review, 8*(1), 80–95.

Wilson, D.B., Mitchell, O., & MacKenzie, D.L. (2006). A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology, 2*(4), 459–487.

Yeaton, W., & Camberg, L. (1997). *Program evaluation for managers: A primer.* Boston: Management Decision and Research Center and Office of Research and Development, Dept. of Veterans Affairs. *Available at* http://www.hsrd.research.va.gov /publications/internal/ProgEval-Primer.pdf

Yu, J., Clark, L.P., Chandra, L., Dias, A., & Lai, T.F. (2009). Reducing cultural barriers to substance abuse treatment among Asian Americans: A case study in New York City. *Journal of Substance Abuse Treatment, 37*(4), 398–406.

Zarkin, G.A., Cowell, A.J., Hicks, K.A., Mills, M.J., Belenko, S., Dunlap, L.J., & Keyes, V. (2012, November 5). Lifetime benefits and costs of diverting substance-abusing offenders from state prison. *Crime & Delinquency: Online.* doi:10.1177/0011128712461904