

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

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

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

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

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

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

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

ARTICLE SUMMARIES

WHAT IS PERFORMANCE MEASUREMENT?

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

MEASURING DRUG COURT PERFORMANCE

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

CONCLUSION

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

INTRODUCTION

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

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

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

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

WHAT IS PERFORMANCE MEASUREMENT?

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

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

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

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

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

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

make sense to include violent offenders in the comparison group.

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

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

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

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

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

What Makes Performance Measurement Valuable?

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

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

MEASURING DRUG COURT PERFORMANCE

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

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

Retention

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

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

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

Sobriety

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

Recidivism

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

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

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

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

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

Units of Service

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

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

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

Client-Level and Program-Level Variables

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

Some Caveats

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

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

CONCLUSION

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

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

REFERENCES

- Bachman, R., & R.K. Schutt. (2003). *The practice of research in criminology and criminal justice* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.
- Brewster, M.P. (2001). An evaluation of the Chester (PA) drug court program. *Journal of Drug Issues*, 31, 177-206.
- Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.
- Epstein, P.D., Coates, P.M., Wray, L.D., & Swain, D. (2005). *Results that matter: Improving communities by engaging citizens, measuring performance and getting things done*. San Francisco: Jossey-Bass Publishing.
- Fitzgerald, J.D. & Cox, S.M. (1994). *Research methods in criminal justice: An introduction*. Chicago: Nelson-Hall Publishing.
- Johnson, C.M., & Wallace, S. (2004). Critical elements to consider for methodologically sound impact evaluations of drug court programs. *Drug Court Review IV*(2), 35-48.
- King, G., Keohane, R.O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Marlowe, D.B., Festinger, D.S., & Lee, P.A. (2004). The judge is a key component of drug court. *Drug Court Review IV*(2), 1-34.

- Roman, J., Townsend, W., & Bhati, A. (2003, July). *National estimates of drug court recidivism rates*. Washington, DC: National Institute of Justice, U.S. Department of Justice.
- U.S. Government Accountability Office. (2005, February). *Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes*. Report to congressional committees. Washington, DC: Author.